

THIS WEEK

EDITORIALS

WORLD VIEW Debate on the impacts of e-cigarettes is smokescreen **p.349**

OPIOIDS Yeast could make painkillers without poppies **p.350**



METEORITE Collision scars reveal high-impact space journey **p.350**

Diplomatic service

Government science advisers are unlikely to be specialists on the subject of a crisis, but they are key to bringing together relevant experts and disseminating the information clearly and accurately.

If all political careers end in failure, then what does that say about the scientific advice those politicians receive? Academic wisdom on tap is becoming more common in policy-making. Almost 60 years after the launch of the first Soviet satellite Sputnik startled US President Dwight Eisenhower into appointing James Killian as his nation's first scientific adviser, Killian's modern equivalents from around the world will meet in New Zealand this week for their own conference. They will, presumably, keep their mobile phones switched on.

Researchers and politicians want different things from these people. To working scientists, an adviser is one of their own who has the ear of government and has the chance to talk up the importance of science and to protect national investment in research. Politicians, however, want the stamp of peer-reviewed approval for their policies. Sometimes, they even want those policies to be based on the peer-reviewed evidence. Incidentally, none of these functions was a priority for Eisenhower, who wanted to harness domestic science to improve US economic and military prospects (see *Nature* **488**, 559; 2012).

Science advice to governments, either through the formal mechanism popular in the United States, Britain and some other European countries, or through more ad hoc systems, often takes on an emergency-response and crisis-management role. This topic is a special focus for the New Zealand meeting. On page 360 of this issue, *Nature* highlights and analyses three examples in which scientists were at the centre of a national crisis. A volcanic eruption in Iceland in 2010 grounded flights across Europe and saw the crude output of atmospheric models discussed on the front pages of tabloid newspapers. And just as that emergency started to ease, experts in the United States had to wrestle

with the political, environmental and economic fallout of the fatal explosion of the BP Deepwater Horizon drilling rig and the massive oil leak into the surrounding Gulf of Mexico. Then, a year later, dozens of people in Germany were killed and hundreds more hospitalized in Europe's worst recorded outbreak of *Escherichia coli* poisoning.

Science and science advice received mixed reviews in each of these incidents. The research got there in the end: the oil spill was contained and quantified, and the source of the *E. coli* identified. Should another volcanic eruption burst through the Iceland ice — and one was threatening to as *Nature* went to press — regulators now have more refined models with which to work out the likely impact. But, too often, the scientific response to a problem is overtaken by events.

A week is a long time in politics, but it is an eternity in a crisis. Britain's response during the 2010 eruption offers a model that could be emulated elsewhere. The country's chief science adviser was able to tap into a previously developed crisis-response strategy and convene a broad panel of scientific and technical experts that included people both inside and outside government. Events still moved faster than the panel did, and key decisions were taken before the committee even met, but once the scientists gathered, they provided important advice that helped the government to decide how to respond to the developing emergency.

Communication, of course, is key. Peter Gluckman, New Zealand's chief science adviser, is hosting this week's conference and wants it to spawn a network of science advisers who can learn from past crises and call on each other for help. "We're a small country. If something happens which is not in our area it would be nice to ring up someone in Europe who has the expertise," he says. Keep those phones on, folks. ■

People power

Climate models must consider how humans are responding to a warming world.

Physics and mathematics can tell us how the Universe began, but as the cosmologist Stephen Hawking noted: "They are not much use in predicting human behaviour because there are far too many equations to solve."

The motives, needs and desires that drive human action have long resisted rational analysis. From the volatility of the stock market to fads and fashions that flare brightly and then vanish, the ability of individuals to act unpredictably has undermined attempts to model their behaviour with any level of precision.

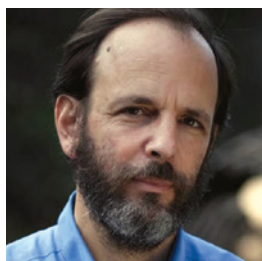
The science-fiction writer Isaac Asimov had the right idea. If one considers a sufficiently large population of people, he wrote, then

just as the mass movement of a gas can be inferred through simple calculations — whatever the individual molecules might do — so too can the future actions of a large population.

Asimov called his fictional science of predicting people's behaviour psychohistory. He used it as a central plank of his classic *Foundation* series of books. The predictions of psychohistory were more than a model, they were a set of instructions for how future societies must respond to a predictable crisis they helped to create.

In a Comment on page 365, Paul I. Palmer and Matthew J. Smith call for human adaptation to climate change to be modelled to help avert a real-life predictable crisis. Existing models of the planet's changing climate are insufficient, they argue, because they leave out the people. Omitting human behaviour from these mathematical studies, they write, is like "designing a bridge without accounting for traffic".

Societies will be different in a warmer world, they point out, and we should understand how this will unfold. It is, in essence, another feedback in the climate system, and one that should be quantified and accounted for. Perhaps another seven billion equations will need to be added to the mix. ■



Allow use of electronic cigarettes to assess risk

Monitoring the outcomes of incentivized e-cigarette use, not endless research, will be the key to sensible regulation, says Daniel Sarewitz.

Electronic cigarettes are growing rapidly in popularity. In the United States they remain unregulated, but 8 August marked the end of the public comment period on a proposal by the US Food and Drug Administration (FDA) that would bring e-cigarettes under its authority. Now the FDA must act.

That action must be based on science. The FDA's background document explains: "We do not currently have sufficient data about these products to determine what effects e-cigarettes have on the public health." Only if e-cigarettes are deemed a tobacco product, as under the proposal, can the agency begin to collect the data that can permit it to "account for the net public health impacts". On the basis of this assessment, the FDA can decide how to craft regulations to protect and improve public health.

When it comes to the direct impact of these devices on public health, the tenor of the FDA's language, and of comments from many relevant organizations, is precautionary. The American Cancer Society (ACS) says: "Until electronic cigarettes are scientifically proven to be safe and effective, ACS will support the regulation of e-cigarettes and laws that treat them like all other tobacco products." The Forum of International Respiratory Societies goes further, saying that the risks of e-cigarettes have not been adequately studied and as a precaution, such devices "should be restricted or banned until more information about their safety is available". A group of 29 state attorneys general sent a 33-page letter to the FDA arguing that "e-cigarettes contain and deliver nicotine — a well-recognized addictive chemical — in amounts comparable to traditional cigarettes. Accordingly, e-cigarettes should be assumed to be both harmful and addictive."

Bollocks. Let's do a thought experiment. Imagine that every smoker in the United States changed to e-cigarettes. What would be the consequences? An e-cigarette, in essence, allows you to be addicted to nicotine (which is not carcinogenic), and to enjoy the tactile pleasures of smoking without exposing yourself to the 60 or more cancer-causing agents, or to most of the hundreds of other toxic chemicals, that are released from burning tobacco. If all US smokers 'vaped' (the verb coined to distinguish inhaling e-cigarette vapours from inhaling tobacco smoke) instead of smoked, about 480,000 deaths might eventually be avoided per year. We may never approach such a full transition, but the point is that the causal relationship between inhaling tobacco smoke and dying from cancer and other diseases is very robust.

How many people would e-cigarettes kill instead? Evidence of the effects of widespread vaping is limited and contradictory. Some studies show that e-cigarettes can wean people from smoking, others suggest that the effect is, at best,

modest and short-lived. Still others suggest that anything that makes vaping seem desirable might coax non-smokers to smoke.

Unanticipated potential risks are being discovered and debated. Studies have shown that e-cigarette vapour includes various fine particulates, some of which are toxic; other research indicates that exposure levels are too low to be dangerous. More science will expand the evidence and the potential risks, but as complexities and questions emerge it will also increase, rather than reduce, the contradictions and uncertainties. The extraordinary difficulty of demonstrating the benefits of salt reduction, mammography, or various diets, for example, ought to serve as cautionary lessons. Given the millions who will die from smoking in the near future, does it make sense to spend years discovering, characterizing and debating ancillary risks of vaping that

are almost certainly less serious than the known risks of smoking as a precondition for responsible policy-making? This is precaution?

E-cigarettes must be regulated. Ingredients should be labelled. No responsible voices would allow them to be sold to children. Such requirements are already in force in the European Union. The more important question is whether regulation should be driven by the risks of e-cigarette use, or by the risks of not using them. The former promises endless research, uncertainty, and debate; the latter may offer a technological short-cut to solving one of the world's most serious public-health problems.

No one knows to what extent vaping will displace smoking, but the sure way never to find out is to make policies hostage to endless studies on population-wide risks. Instead we should

test the effectiveness of policies, perhaps in limited jurisdictions, that encourage vaping among smokers and potential smokers. Keep the tax burden, and thus cost, low relative to cigarettes. Allow advertisements. With George Clooney. Continue to allow vaping in bars, restaurants and workplaces. Make smoking uncool, expensive and stupid, and vaping cool and smart. If people must get addicted to something, let them get addicted to a thing that does not give them or their families cancer. And carefully monitor the outcomes.

As research for this column, I tried vaping. The taste was perfectly pleasant, and my office colleagues said that the white-ish clouds I exhaled had no smell. I am not a smoker, so I cannot rate how the overall experience compares to cigarettes, and I have not vaped enough to recognize any physiological response. But then again, I am already addicted to the pleasures of caffeine, and one costly addiction is enough. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: daniel.sarewitz@asu.edu

**MAKE SMOKING
UNCOOL,
EXPENSIVE
AND STUPID,
AND VAPING
COOL AND
SMART.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/fyqhwr

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

BIOTECHNOLOGY

Yeast turned into opioid-makers

Baker's yeast can be made to produce morphine, codeine and other pain medicines at high levels.

Opioid production is subject to the vagaries of the opium poppy supply chain, so Christina Smolke and her colleagues at Stanford University in California wanted to find alternative ways of making the drugs. They engineered the yeast (*Saccharomyces cerevisiae*) to express genes from the poppy (*Papaver somniferum*) and the bacterium *Pseudomonas putida* M10, then cultured it with thebaine, an opioid intermediary molecule extracted from the poppy. The yeast synthesized high enough levels of several natural and semisynthetic opioids to make the method potentially useful to the pharmaceutical industry, the authors say.

The next step, they add, is to engineer yeast to make these painkillers from simple sugars, eliminating the need for poppies altogether.

Nature Chem. Biol. <http://dx.doi.org/10.1038/nchembio.1613> (2014)

ASTRONOMY

Collision history written in rock

Meteorites recovered in California have yielded details about their collision-filled journey from the Solar System's asteroid belt.

The fragments (pictured) originated from a meteoroid whose fiery descent lit up the night sky over San Francisco in 2012. Peter



Jenniskens of NASA's Ames Research Center in Moffett Field, California, and his colleagues analysed the six fragments and concluded that the meteoroid is one of several thought to have come from a parent asteroid that shattered in a collision 470 million years ago. This meteoroid broke

apart again in collisions 9 million and 4 million years ago. Measurements of accumulated radiation in the rock suggest that it suffered a final collision within the past

100,000 years.

The team also detected traces of carbon-rich organic molecules that somehow managed to survive.

Meteorit. Planet. Sci. 49, 1388–1425 (2014)

MICROBIOLOGY

Liquid layer for lung defence

Inhaling bacteria increases the production of mucus and liquid that trap and kill microbes in the airways.

A team led by Juan Iwanowski at the University of

water, *Acropora* coral larvae spent more than 85% of their time in water from the protected areas. The organisms also showed a similar preference for clean water containing chemical cues from a variety of corals, whereas seaweed cues decreased this preference.

Conservationists will probably need to boost coral-attracting chemicals when rebuilding seaweed-choked reefs, the authors suggest.

Science 345, 892–897 (2014)



MARINE ECOLOGY

Sick reefs repel young coral

Coral larvae actively avoid the smell of degraded marine ecosystems — potentially impeding efforts to rebuild damaged reefs.

Mark Hay at the Georgia Institute of Technology in Atlanta and his team studied coral reefs near Fiji (pictured). They focused on the behaviour of coral larvae in water from healthy, protected reefs and from reefs overgrown with seaweed.

Given the choice between the two streams of

Saskatchewan in Saskatoon, Canada, developed an imaging technique to visualize the depth of the protective layer that lines the airways. The researchers found that when the bacterium *Pseudomonas aeruginosa* was introduced into isolated pig tracheas, the airways secreted more of this film than unexposed tracheas.

The results could explain why patients with cystic fibrosis, who do not produce this layer properly, are prone to respiratory infections and inflammation. Boosting this previously unknown part of the immune response in these

JOÃO PAULO KRAJEWSKI

P. JENNISKENS/NASA ARC

patients could improve their lung function, the authors say. *Proc. Natl Acad. Sci. USA* <http://doi.org/vbc> (2014)

NEUROSCIENCE

Light signals boost stroke recovery

Mice that have had a stroke regain motor function after specific brain neurons are stimulated by light.

Using a technique called optogenetics, Michelle Cheng, Gary Steinberg and their colleagues at Stanford University in California studied mice that were genetically engineered to express a light-sensitive protein in a group of neurons in the brain's motor cortex.

The team induced a stroke on one side of this area and then activated the engineered neurons by switching on a small, implanted light probe. This stimulation increased blood flow and the levels of various growth factors in the opposite, uninjured cortex. Moreover, these animals gained more weight and walked faster and farther than unstimulated stroke mice.

Further study of the mechanisms of this recovery could lead to new drug targets or therapy techniques for stroke, the authors say. *Proc. Natl Acad. Sci. USA* <http://doi.org/vbj> (2014)

LIMNOLOGY

Earth's lakes added up

The world is home to 117 million lakes, according to a satellite-based survey.

Lakes are important to the planet's carbon cycle, but researchers have struggled to estimate the number that exist, in part because small lakes have been difficult to count. A team led by Charles Verpoorter of the University of Lille Nord in France used high-resolution imagery from the Landsat satellite to catalogue all lakes larger than 2,000 square metres.

Unlike earlier estimates, the total count is dominated by large and medium-sized lakes, not small ones. Together, the lakes cover about 5 million square kilometres, or 3.7% of Earth's land not covered by ice. *Geophys. Res. Lett.* <http://doi.org/t9v> (2014)

Virology

Polio killed the vaccinated

The poliovirus strain that caused an outbreak in the Republic of the Congo in 2010 is able to resist the immune responses generated by a commonly used vaccine. The finding could explain why the outbreak, which killed nearly half of the 445 people infected, was so severe.

Christian Drosten at the University of Bonn Medical Centre in Germany and his colleagues analysed the virus strain responsible for the outbreak. They found a combination of two mutations, both in the proteins of the strain's 'coat', which make it harder for certain antibodies to stick to the virus. People who died in this outbreak had been vaccinated in the past, but people who were freshly re-vaccinated could fight off the virus.

The authors warn that other resistant strains could emerge, even as the world is close to eradicating the disease. *Proc. Natl Acad. Sci. USA* <http://doi.org/vbg> (2014)

ENVIRONMENTAL MICROBIOLOGY

Algal boom and bust tracked

Viruses that infect and kill algae could influence the ocean carbon cycle.

Ilan Koren and Assaf Vardi at the Weizmann Institute of Science in Rehovot, Israel, and their colleagues used satellite imagery to measure the atmospheric carbon absorbed by a roughly 1,000-square-kilometre algal bloom during its 25-day life cycle in the

SOCIAL SELECTION

Popular articles on social media

The cost of misconduct

A report detailing the supposedly meagre costs of scientific misconduct has set off an online discussion about the real toll of shoddy science. Using the records of the US government's Office of Research Integrity, researchers led by Ferric Fang at the University of Washington in Seattle collected studies funded by the National Institutes of Health (NIH) that had been retracted because of misconduct from 1992 to 2012. All told, the NIH had spent about US\$58 million on these projects, less than 1% of its total budget over this time. On Google+, mathematician Joerg Fliege at the University of Southampton, UK, did what he does best: the maths. "Multiply this figure with 100 if you believe that only 1% of misconducts get detected," he posted. "Still not much of an expense, in the overall scheme of things."

eLife 3, e02956 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/tuaco5

North Atlantic. They found that the algae converted around 22,000 tonnes of atmospheric carbon into organic carbon — about as much as a rainforest of equivalent size — before viruses caused the bloom to collapse.

Two-thirds of this carbon was released back into the atmosphere within a week of the bloom's collapse. The rest could have been transported deep into the ocean, as infected algae sank to the ocean floor, the authors suggest.

Curr. Biol. <http://doi.org/vbx> (2014)

PLANT SCIENCES

Plants drink mineral water

Thirsty plants can extract water from the crystalline structure of gypsum, a rock-forming mineral found in soil on Earth and Mars.

Some plants grow on gypsum outcrops and remain active even during dry summer months, despite having shallow



roots that cannot reach the water table. Sara Palacio of the Pyrenean Institute of Ecology in Jaca, Spain, and her colleagues compared the isotopic composition of sap from one such plant, called *Helianthemum squamatum* (pictured), with gypsum crystallization water and water found free in the soil. The team found that up to 90% of the plant's summer water supply came from gypsum.

The study has implications for the search for life in extreme environments on this planet and others.

Nature Commun 5, 4660 (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Science adviser

The post of chief scientific adviser to the European Commission came under attack again on 19 August when nearly two dozen non-governmental organizations called on the incoming president of the commission to scrap the job. In an open letter to Jean-Claude Juncker, groups including Friends of the Earth Europe added their weight to an existing campaign to abolish the science role. The letter argues that the position “concentrates too much influence in one person”. The mandate of the commission’s current chief scientific adviser, Anne Glover, is scheduled to end later this year.

Trial enrolment

The US Food and Drug Administration published a plan on 20 August to ensure that women and racial and ethnic minority groups are adequately included in clinical trials. The action plan aims to collect data on population subgroups in trials, to encourage and enable more women and minority groups to enrol, and to make the demographic make-up of trials more transparent. It will be implemented in stages over the next five years. The agency also published guidance on evaluating how medical devices might function differently in men and women.

EVENTS

Icelandic volcano

After a reported eruption on 23 August turned out to be a false alarm, Iceland’s Bárðarbunga volcano continued to rumble deep underground. Since seismic activity began on 16 August, thousands of earthquakes have shaken the ground north and

east of the volcano. They show where magma is squirting up from below and forming a freshly cooled sheet of rock, or dyke, a few kilometres deep. As of 25 August, the dyke was thought to be nearly 35 kilometres long and to contain 300 million cubic metres of magma. Aviation authorities remained on alert in case an eruption spewed ash into the air. See go.nature.com/iidaau for more.

Ebola in fifth nation

The Democratic Republic of Congo became the fifth African nation to confirm cases of Ebola, on 24 August. On the same day, a Senegalese epidemiologist was reported as the first person working for the World Health Organization (WHO) to

contract the disease. On 22 August, the WHO said that the official count of 2,615 cases and 1,427 deaths probably underestimates the true size of the epidemic. It blamed community resistance to reporting cases and a lack of adequate treatment facilities. The agency thinks that the epidemic could last for another 9 months. See page 355 for more.

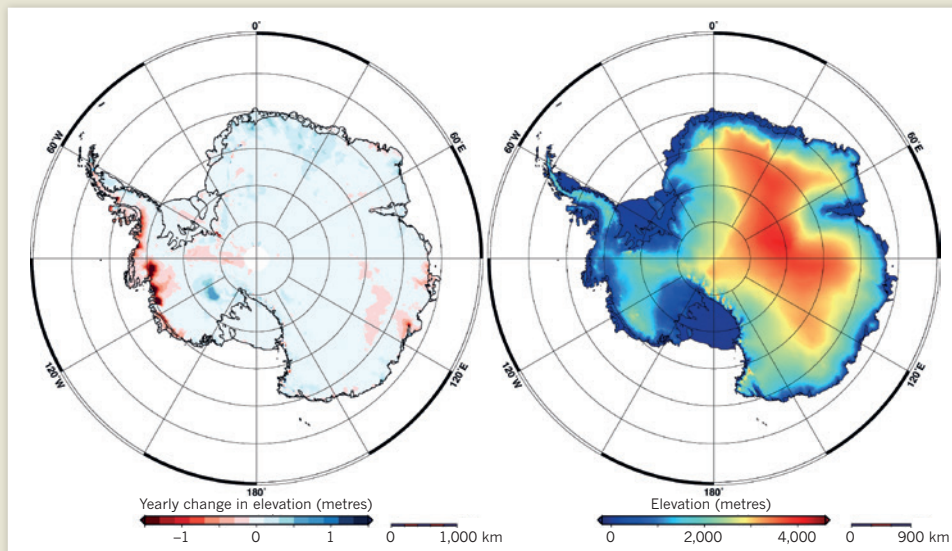
Wonky orbit

The latest pair of satellites in Europe’s Galileo global positioning system went into the wrong orbit after a botched launch from French Guiana on 22 August. The satellites were supposed to adopt a circular orbit, but are now travelling on an elliptical path and at a lower altitude than

planned. Arianespace, the French company responsible for the launch, said that it is investigating whether the problem was due to the Soyuz launch rocket. The satellites are the fifth and sixth out of a total constellation of 30 Galileo spacecraft.

Drone lawsuit

An association of 188 US universities is taking the government to court in an attempt to overturn rules that restrict researchers’ use of drones. On 22 August, the Council on Governmental Relations in Washington DC filed a complaint in the federal appeals court, arguing that restrictions on the ‘commercial’ use of drones — imposed by the US Federal Aviation Administration



Satellites pinpoint ice loss

The massive ice sheets in Antarctica (pictured) and Greenland are together shrinking at a rate of 500 cubic kilometres per year — the fastest pace since satellite altimetry began 20 years ago. The European Space Agency’s CryoSat-2 probe collected precise elevation data for both ice sheets (Antarctica’s shown on right). The

greatest ice loss (shown on left in red) between 2011 and 2014 occurred at Pine Island glacier in western Antarctica and at Jakobshavn glacier in Greenland. The findings were reported on 20 August by researchers from the Alfred Wegener Institute in Potsdam, Germany (V. Helm *et al. Cryosphere* 8, 1539–1559; 2014).

(FAA) in June — unfairly hamper the application of drones in science (see also *Nature* **512**, 231; 2014). The FAA says that commercial use includes research activities at private universities.

PEOPLE

Nuclear spy

A former employee of the Los Alamos National Laboratory in New Mexico was convicted on 20 August of conspiring with her husband to sell US nuclear secrets to Venezuela. Marjorie Roxby Mascheroni, who was a writer and editor at the lab between 1981 and 2010, was handed a 366-day prison sentence. The charges arose from the interactions of her husband, formerly a physicist at the lab, with an undercover FBI agent posing as a Venezuelan government official. Pedro Leonardo Mascheroni, who awaits sentencing, claims that his dealings were part of an attempt to fund construction of a laser for nuclear fusion.

Minister dismissed

Iran's parliament voted on 20 August to dismiss the country's science and technology minister, Reza Faraji-Dana (**pictured**), for attempting to liberalize universities and allegedly politicizing Iran's academic environment. Faraji-Dana, an



electrical engineer and former chancellor of the University of Tehran, had joined the cabinet of reformist president Hassan Rouhani last year. He upset hardline parliamentarians when he tried to allow activist students and professors back onto campus after they had been banned by Rouhani's conservative predecessor Mahmoud Ahmadinejad following anti-government unrest in 2009.

RESEARCH

MERS model

Marmosets are the best animal model for Middle Eastern Respiratory Syndrome (MERS). Research on the MERS coronavirus was hindered by the lack of an animal model that showed the same respiratory symptoms as humans when infected with the virus. In two studies published on 21 August,

researchers from the US National Institute of Allergy and Infectious Diseases showed that the virus could infect marmosets, and that the animals' symptoms mimic the severe pneumonia seen in humans (D. Falzarano *et al.* *PLoS Pathog.* **10**, e1004250 (2014); N. van Doremalen *et al.* *J. Virol.* **88**, 9220–9232 (2014)).

African farming

The Alliance for a Green Revolution in Africa (AGRA), a science-based non-governmental organization in Nairobi has helped 1.7 million African farmers to rejuvenate 1.6 million hectares of land, and to double, or even triple, crop yields over the past five years through its Soil Health Programme. AGRA reported the results on 22 August. The programme tests and teaches techniques to improve soil fertility and makes chemical fertilizers more affordable for poor farmers. Depleted soils cost African farmers US\$4 billion a year in lost productivity.

Stem-cell go-ahead

Regulators in the United States have cleared the way for a clinical trial of a prospective stem-cell-derived treatment for type 1 diabetes. On 19 August, ViaCyte of San Diego, California, said that the Food and Drug

COMING UP

2–4 SEPTEMBER

Agricultural scientists meet at the Africa Green Revolution Forum in Addis Ababa to discuss how to boost agricultural productivity in the developing world. www.agrforum.com

5–9 SEPTEMBER

Researchers meet at the Interscience Conference on Antimicrobial Agents and Chemotherapy in Washington DC to discuss infectious-disease control and prevention on a global scale. go.nature.com/8cjcaa

Administration has given it permission for a phase I/II clinical trial of a product that consists of pancreatic precursor cells packaged in a mesh pouch. ViaCyte's treatment could become one of only a handful of human embryonic-stem-cell-derived products to be trialled in people, and will be an important test for the effectiveness of California's state stem-cell institute, which provided funding to develop the product.

BUSINESS

Pharma takeover

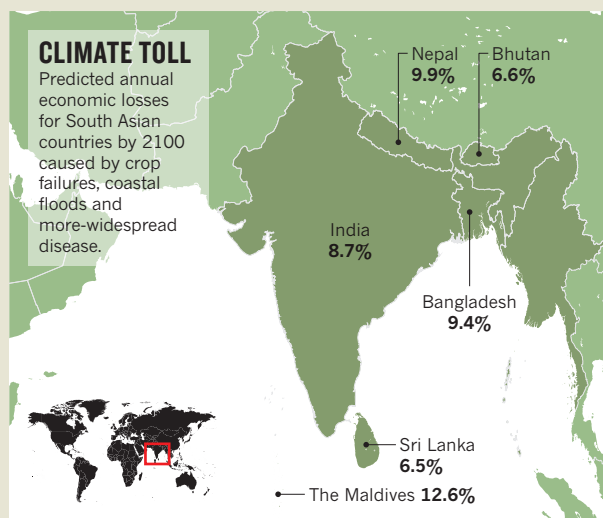
Swiss pharmaceutical giant Roche announced on 24 August that it will be purchasing the biotechnology firm InterMune for US \$8.3 billion. Headquartered in Brisbane, California, InterMune owns a drug called pirfenidone for treatment of idiopathic pulmonary fibrosis. The drug is sold in Europe and Canada, but has yet to gain approval in the United States.

► **NATURE.COM**

For daily news updates see: www.nature.com/news

TREND WATCH

Climate change will cause the collective economy of the six countries of South Asia to lose an average 1.8% of its annual gross domestic product by 2050, rising to 8.8% by 2100. The data, from the Asian Development Bank and published on 19 August, show that the Maldives and Nepal will be hardest hit (see map). The forecast assumes a 4.6°C rise in global temperatures. South Asia will need to spend at least US\$73 billion a year between now and 2100 to adapt to climate change, the report says.



NEWS IN FOCUS

BIOTECHNOLOGY Brazil mulls genetically modified eucalyptus **p.357**

CONSERVATION Whale-watching threat to sea mammal populations **p.358**

CLIMATE CHANGE Himalayan plants scaling new heights **p.359**



SCIENCE ADVICE What happens when scientists face a national crisis **p.360**

JOHN MOORE/GETTY IMAGES



A health-care worker dons protective gear before entering an Ebola treatment centre near Monrovia, Liberia.

INFECTIOUS DISEASES

World struggles to stop Ebola

Greater international assistance is needed to quell the epidemic, say health officials.

BY ERIKA CHECK HAYDEN

Dan Kelly felt as if he were entering a war zone when he arrived at Connaught Hospital on 19 August. His friend Modupeh Cole, the physician in charge of the Ebola isolation ward at the hospital in Freetown, Sierra Leone, had died six days earlier. Marta Lado, a doctor from Spain, was caring for the ward's ten patients. "She was mopping the floors herself," says Kelly, an infectious-disease physician and co-founder of the Wellbody Alliance,

a non-profit health-care organization in Sierra Leone. The international aid group Médecins Sans Frontières has called the world's response to the Ebola epidemic in West Africa "dangerously inadequate". As Kelly travelled around Freetown, noticing closed clinics and health-care workers without adequate protective training and equipment, he had to agree.

"We're lacking the kind of passionate humanitarian response that we've seen for disasters like the Haiti earthquake or the Haiyan typhoon" in the Philippines, says Kelly, who is raising

money through the University of California, San Francisco, to teach infection-control practices to health-care workers fighting the outbreak. More than 240 health-care workers have contracted Ebola in the current episode, including a Senegalese epidemiologist who is the first person deployed by the World Health Organization (WHO) to become infected with the disease. "We're around rock bottom," says Kelly.

Weeks after the WHO declared the Ebola outbreak a public-health emergency of international concern and the World Bank ►

► pledged to provide up to US\$200 million in aid, enough resources to end the crisis have yet to materialize. The WHO reports that the Ebola outbreak has so far claimed 1,427 lives in four West African nations — more than all previous recorded outbreaks combined, and probably an underestimate (see 'Out of control'). It says that the epidemic will last for another 6–9 months, and the scientists and doctors who are fighting it agree. "I can comfortably say that we will be in this for the next 4 to 9 months, not including the follow-up that will be required to deal with the post-traumatic effects," says virologist Joseph Fair, a special adviser to the ministry of health in Sierra Leone.

To end the crisis, developed countries and aid groups need to send more people with expertise in treating infectious diseases, public-health experts say. In Monrovia, Liberia's capital, only three Ebola treatment facilities were operating as of 22 August; the WHO says that, in the coming weeks, facilities will be needed to treat another 500 patients. The US Centers for Disease Control and Prevention in Atlanta, Georgia, has sent 60 people to West Africa to help with surveillance, disease tracking and education, but they do not treat patients. Health-care workers and experts are badly needed from other Western governments, says Sophie Delaunay, executive director of Médecins Sans Frontières in New York. Many institutions and aid groups are even removing doctors from the region

out of concern for the physicians' safety.

Affected countries must also do a better job of dealing with patients and educating the public about the disease, say doctors and scientists. Insensitivity on the part of government officials has sparked uprisings, including a riot in a quarantined neighbourhood of Monrovia on 20 August, says Bailor Barrie, co-founder of the Wellbody Alliance.

Officials are likely to make more progress fighting Ebola's spread if they explain their actions and show concern for those affected, he adds, citing his own experience in a village in Kono, Sierra Leone, earlier this month. After a villager there tested positive for Ebola, police put two houses under quarantine — and six people fled. Barrie, along with community leaders and public-health workers, then visited the remaining 30 residents to explain the reason for the quarantine and to answer questions. One man said that he needed someone to go and buy cigarettes, and another needed palm wine; health officials tasked a worker with making these daily deliveries. The inhabitants stayed in quarantine for the full 21-day isolation period, even as three of them developed Ebola-like symptoms. Diagnostic tests were negative, and no one in the houses came down with the virus.

"We're at the point where almost anything we do to turn this around will make things better."

"When you don't engage people, and just tell them what is right without listening to them, they will not comply with your orders," says Barrie. "That's why the initial response in some countries was full of denial, fear and terror — everyone was afraid."

Fair says that governments also need to undertake widespread, country-wide campaigns, "blasting all forms of media — billboards, radio jingles, T-shirts, pens ... basically, inundating the public with so much information that they can't not hear it and remember it".

But, ultimately, the most effective tactic to fight Ebola's spread is to train West Africans in basic public-health measures, says Daniel Bausch, a physician at Tulane University in New Orleans, Louisiana, who has treated the disease in Guinea and Sierra Leone. "I'm not naive enough to think it's an easy thing to do, but it's the only thing that's actually going to work," he says.

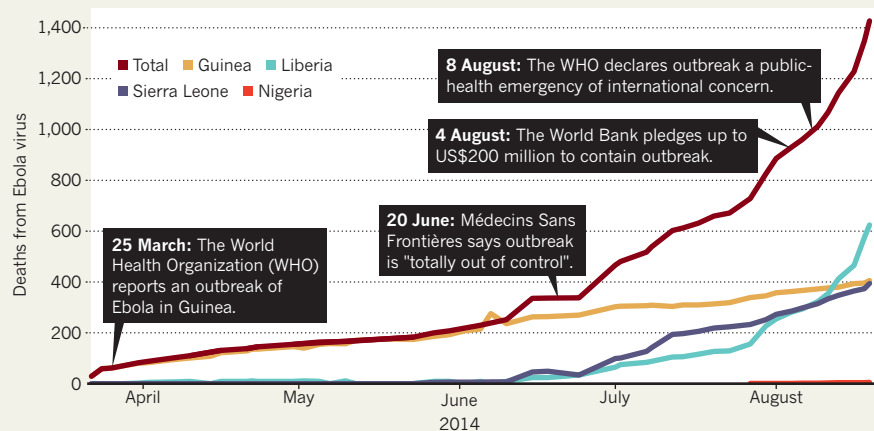
The experience of groups such as Last Mile Health, a non-profit organization that operates in Boston, Massachusetts, and in Monrovia, shows that such training can work. Last Mile Health has deployed 150 community health workers to educate rural Liberians about Ebola. It has also worked to provide basic materials to care for patients. Rajesh Panjabi, the organization's chief executive, says that in mid-August, nurses at the Martha Tubman Memorial Hospital in Zwedru — which serves 130,000 people in rural eastern Liberia — were preparing to strike over a lack of protective equipment. Last Mile Health provided gloves, masks, gowns and training in how to use them, and the nurses continued working. (Liberia's ministry of health had estimated earlier this month that the country needs more than 451,360 pairs of medical gloves.) "This is the type of impact you can have through very basic activities, but it takes sustained response and investment," says Panjabi.

There are other hopeful signs. At Connaught Hospital, nurses from countries including Sierra Leone have joined Lado, the physician leading the clinical effort of the King's Sierra Leone Partnership, an initiative of the King's Centre for Global Health in London. Two doctors and another nurse from the partnership will arrive at Connaught in the coming weeks. "We're at the point where almost anything we do to turn this thing around will make things better," says Kelly. "It's got to be one hospital at a time." ■

SOURCE: WHO

OUT OF CONTROL

The death toll from Ebola virus in West Africa continues to rise. Infectious-disease experts say that more health-care workers are needed to contain the outbreak.



MORE ONLINE

EXPLAINER



How Argentina's default affects research
go.nature.com/ks8fzs

MORE NEWS

- Methane plumes from Atlantic sea floor go.nature.com/5llkts
- Chemical signature found of primordial monster star go.nature.com/s24sa9
- Ionic liquids destroy bacterial biofilms go.nature.com/px8pg5

NATURE PODCAST



Teaching fish how to walk; quantum imaging without light; and the microbes that make cheese nature.com/nature/podcast



Eucalyptus plantations near São Paulo in Brazil.

BIOTECHNOLOGY

Brazil considers transgenic trees

Genetically modified eucalyptus could be a global test case.

BY HEIDI LEDFORD

Viewed from above, Brazil's orderly eucalyptus plantations offer a stark contrast to the hurly-burly of surrounding native forests. The trees, lined up like regiments of soldiers on 3.5 million hectares around the country, have been bred over decades to grow quickly.

On 4 September, a public hearing will consider bringing an even more vigorous recruit into the ranks: genetically engineered eucalyptus that produces around 20% more wood than conventional trees and is ready for harvest in five and a half years instead of seven. Brazilian regulators are evaluating the trees for commercial release; a decision could come as early as the end of this year.

Researchers, businesses and activists are watching closely. Eucalyptus (*Eucalyptus* spp.) — native to Australia — is grown on about 20 million hectares throughout the tropics and subtropics, and approval of the genetically engineered trees in Brazil could encourage their adoption elsewhere. "It would have ripple effects worldwide," says Zander Myburg, who studies the genetics of forest trees at the University of Pretoria in South Africa. "Everybody will pay attention."

So far, no genetically modified tree from a major commercial species has been deployed on

a large scale. The ubiquity of eucalyptus makes Brazil's decision on the modified trees a special concern to environmental activists who oppose the use of genetically modified crops.

"They have become the target of very intensive and emotionally charged debate particularly among the NGOs and nature constituencies," says Walter Kollert, a forestry officer with the Food and Agriculture Organization of the United Nations in Rome.

A consortium of activists opposed to the plan intends to present a letter at the 4 September meeting, urging Brazil's National Technical Biosafety Commission to reject the trees. In all, 259 organizations — 106 of them from Latin America — have signed the letter, which expresses concern that the trees pose risks to the environment and will encourage the expansion of plantations.

The trees were developed by FuturaGene, a biotechnology firm in Rehovot, Israel, that was spun out of the Hebrew University in Jerusalem in 1993. The company found that certain proteins accelerate plant growth by facilitating cell-wall expansion. FuturaGene inserted into eucalyptus a gene that encodes one such protein from thale cress (*Arabidopsis thaliana*), a common laboratory plant. In 2010, the firm was bought by Suzano Pulp and Paper of São Paulo, Brazil, one of the world's largest producers of eucalyptus pulp.

FuturaGene's chief executive Stanley Hirsch is quick to point out the environmental benefits of his company's creation. The tree's speedy growth boosts absorption of carbon dioxide from the air by about 12%, he says, aiding in the fight to reduce greenhouse-gas emissions. The genetically modified trees may also require less land to produce the same amount of wood, reducing the conversion of natural forest into plantations.

Hirsch says that the company has tried to avoid public-relations mistakes made by agricultural biotechnology companies in the past: rather than shun activists, he has invited them to tour the company's field-trial sites. "Some of them were so surprised," he says. "They said, 'Wow, these look just like normal trees.'"

Hirsch's pitch has not convinced everyone. Anne Petermann, executive director of the non-profit organization Global Justice Ecology Project in Buffalo, New York, says that FuturaGene is trying to stave off opposition by 'greenwashing' its product. Faster-growing trees require more water and extract more nutrients from the soil, she adds, and they will only add to the economic incentive to seed more plantations.

Genetically engineered trees do pose some biosafety issues that do not apply to agricultural crops such as maize (corn) or soya, notes forest geneticist Steven Strauss of Oregon State University in Corvallis. They remain in the environment for years, increasing their potential impact on the plants, animals and soil around them. And trees tend to disperse pollen further than crops nearer the ground do, raising concerns about gene flow to native relatives. But eucalyptus has no native relatives in Brazil and is not particularly invasive in most areas of the country, says Strauss.

FuturaGene says that it identified no major environmental problems in eight years of field trials that collected data on everything from gene flow to leaf-litter decomposition to the composition of honey made by bees that visit the trees. Myburg, who does not work with FuturaGene but is familiar with the company's safety data, says that he found the firm's studies to be well designed and thorough.

While FuturaGene tests the waters in Brazil, a US company awaits a regulatory decision regarding its genetically engineered, freeze-tolerant eucalyptus. In 2008, ArborGen of Ridgeville, South Carolina, petitioned the US Department of Agriculture to allow commercialization of the trees in the southeastern United States. Delays of this length are not uncommon in the US regulatory system, says ArborGen's director of regulatory affairs Leslie Pearson.

For now, just the prospect that the trees might be approved has been enough to rally activists. "The fact that there are now two commercial applications has the movement against genetically modified trees mobilizing quickly in many regions," says Petermann. "We know we're going to be seeing a lot more coming out from this industry." ■

Ecotourism rise hits whales

Desire to observe whales and dolphins up close is affecting animals' behaviour.

BY DANIEL CRESSEY

Boat trips to watch whales and dolphins may increasingly be putting the survival of marine mammals at risk, conservationists have warned.

Research published this year shows that the jaunts can affect cetacean behaviour and stress levels in addition to causing deaths from collisions. But some animals are affected more than others and the long-term effects remain unclear, scientists at the International Marine Conservation Congress (IMCC) in Glasgow, UK, heard last week.

"Whale-watching is traditionally seen as green tourism," says wildlife biologist Leslie New of the US Geological Survey in Laurel, Maryland. "The negative is the potential for disturbance. That disturbance is a worry because we don't want to do 'death by 1,000 cuts'."

The number of people joining trips has expanded hugely since the 1990s, from 4 million in 31 countries in 1991 to 13 million in 119 countries in 2008, the most recent year for which full data are available. In 2008, the International Fund for Animal Welfare, an animal-protection charity in London, estimated the value of the industry at US\$2.1 billion.

Although collisions with boats can hurt the animals, researchers are more concerned about effects such as animals failing to feed or using up energy swimming away from the vessels.

These seemingly small events can add up, studies suggest.

Earlier this year, for example, marine biologist David Lusseau of the University of Aberdeen, UK, and his team showed that minke whales (*Balaenoptera acutorostrata*) in Faxaflói Bay in Iceland responded to whale-watching boats as they do to natural predators, upping their speed and respiring more heavily¹. But whether this was a direct result of the boats is difficult to pin down:

Lusseau, who was not at the meeting, says that soon-to-be-published research by his team shows that behavioural changes are probably not affecting actual numbers of the minke in Faxaflói Bay.

But Lusseau's group has also shown that the bottlenose dolphins (*Tursiops* sp.) in Doubtful Sound, New Zealand, could be driven to extinction in decades². The large number of dolphin-watching trips in the sound is driving the animals away from their preferred areas and forcing them to avoid boats instead of feeding. Dolphin numbers declined from 67 in 1997 to 56 in 2005, the team found.

Several delegates at the IMCC also described the effects on the roughly 70 endangered Irrawaddy dolphins (*Orcaella brevirostris*) living in the Mekong River between Cambodia

and Laos, which are hounded by scores of tourist boats.

Determining which populations are most at risk could help to fix the problem, says Lusseau. He suggests plugging short-term observational data into longer-term population models to tease out whether behavioural changes are temporary or serious long-term threats. There are enough data on species types and locations to assess, at least roughly, where whale-watching should and should not be allowed, he says. But funding and political support are hampering the creation of detailed, localized plans. "There is a lot of lip service being paid to understanding the challenges tourism poses on wildlife, but in practice there is very little financial interest in finding this out," he says.

SHORT-TERM FIX

Guidelines such as specifying minimum distances between animals and boats, speed limits or no-go areas, can help. But codes vary widely: a 2004 study³ found that just 38% were binding; the rest were voluntary. They are also often inadequate. Even with guidelines in place, boats in the dolphin-watching haven of the Bocas del Toro region of Panama hit and killed at least 10 animals in a population of about 250 in 2012 and 2013, according to research presented to the International Whaling Commission in Cambridge, UK, this year.

Greg Kaufman, executive director of the Pacific Whale Foundation — an organization in Hawaii that runs whale-watching and research trips — holds up the Irrawaddy dolphin as a population desperately in need of protection. "They're basically killing these animals one at a time," he says.

But Brian Smith, a zoologist at the Wildlife Conservation Society in New York who has long studied the Irrawaddy group, says that although cetacean tourism is probably stressful for these animals, the main problem is entanglement in fishing nets. And the alternative to fishing for many people in the region is dolphin-watching.

Most of the speakers at the IMCC meeting agreed that more should be done to protect dolphins and whales from tourists. "Although whale-watching is not as bad as whaling," says New, "it might be that last piece that pushes a species over." ■

1. Christiansen, F., Rasmussen, M. H. & Lusseau, D. *J. Exp. Mar. Biol. Ecol.* **459**, 96–104 (2014).
2. Lusseau, D., Slooten, L. & Currey, R. J. C. *Tourism Mar. Environ.* **3**, 173–178 (2006).
3. Garrod, B. & Fennell, D. A. *Ann. Tourism Res.* **31**, 334–352 (2004).



Whale-watching trips often come into very close contact with the animals.

MICHAEL NOLAN/ROBERT HARDING WORLD IMAGERY/CORBIS



Steep mountains in Himachal Pradesh, India, give rise to a rich variety of species.

CLIMATE CHANGE

Himalayan plants seek cooler climes

Race is on to record mountain biodiversity before it is lost.

BY T. V. PADMA IN SHIMLA, INDIA

In India's Western Himalayas, changes in altitude are so dramatic and steep that alluvial grasslands, subtropical forests, conifers and alpine meadows lie stacked almost on top of each other, producing a spectacular range of vegetation. Now, the myriad plants that inhabit these mountains are migrating upwards because of climate change — and some are in danger of being lost before anyone has even recorded their existence.

"Indian scientists have not documented many of the species in the Western Himalayas, so one does not know what species existed in the first place, and where they have shifted to," says Vaneet Jishtu, a botanist at the Himalayan Forest Research Institute in Shimla in the north Indian state of Himachal Pradesh. Jishtu and his team are midway through a five-year project to catalogue the upward migration.

Although melting glaciers are often the focus of climate-change concerns in the Himalayas, the range is also home to one-tenth of the world's known higher-altitude plant and animal species, and half of India's native plant species. Particularly rich in biodiversity are the Western Himalayas that, in India, include the states of Himachal Pradesh, Jammu and Kashmir, Uttarakhand and Sikkim (see 'Peak plants'). There, mountains that rise steeply from 300 metres to more than 6,000 metres have diverse ecosystems and act as a natural barrier to species migration. But data collection on biodiversity in the Himalayas is sporadic — and there is much less

information than from the European Alps or the Andes, says Nakul Chettri, coordinator of the transboundary landscapes programme at the International Centre for Integrated Mountain Development (ICIMOD) in Kathmandu.

The studies that do exist tend to be on well-known species, but give an idea of the extent of the upward shift. Over the past decade, tropical plants such as the nuisance weed *Parthenium hysterophorus* and the water fern *Azolla cristata* have settled in the formerly temperate climate of the Kashmir Himalayas, according to studies by Zafar Reshi, a botanist at the University

of Kashmir in Srinagar in Jammu and Kashmir state (see, for instance, B. Ahad *et al. Am. Fern J.* **102**, 224–227; 2012). He also notes that the flowering and fruiting cycles of pear and apple trees — for which the region is famous — have shifted in line with rising temperatures during the past 50 years, leading to changes in fruit size, colour and taste (B. Basannagari and C. P. Kala *PLoS ONE* **8**, e77976; 2013).

Himalayan blue pines (*Pinus wallichiana*) are on the move too. They are now seen at heights of 4,000 metres, whereas two or three decades ago they grew at altitudes no higher than 3,000 metres, says a team led by Sher Singh Samant, head of the biodiversity and conservation team at the Himachal Pradesh unit of the G. B. Pant Institute of Himalayan Environment and Development. "In the past, heavy snowfall in higher regions prevented the upward shift of species. Now there is less snow and it has started melting faster, bringing about changes in vegetation and alpine meadows," he says. Even species among the glacier-deposited rocks at 4,500–5,500 metres are moving upwards to cooler climes, he notes.

Researchers know much less about — or have not even documented — many of the region's native species, especially ones that inhabit high-altitude transition zones, or ecotones, where one ecosystem gives way to another. Species in these zones are most affected by climatic change, says Chettri: "They either have to adapt by shifting to higher altitudes, or they die."

Such fears prompted the launch in 2012 of Jishtu's five-year project to assess the impact of global warming on transition zones in Himachal Pradesh. "We are still collecting and analysing data," he says. He is also setting up an arboretum in Shimla to conserve Himalayan plants that are endangered or rapidly disappearing.

One problem with existing Himalayan biodiversity assessments is a tendency to limit them to protected areas and certain species, says Chettri. This affects data on animals as well as on plants. He cites a draft ICIMOD analysis of the Eastern Himalayas that noted the cataloguing of biodiversity there is skewed towards charismatic species such as the red panda (*Ailurus fulgens*), with lower vertebrates and invertebrates sidelined. Even preliminary information about algae, fungi, insects, mosses and ferns is still scarce in the Kashmir Himalayas in India, adds Reshi.

The data that do exist can be hard for researchers to find. Past records of the region's biodiversity are scattered, inaccessible or lost in obscure natural history museums, says Reshi. "There is little systematic digitization of the collected data, and access to the data is limited, leading to duplication and misuse of resources and time," adds Chettri.

And time could be in short supply: species can migrate upwards for only so long before they hit inhabitable terrain. The upper Himalayas comprise mostly rocks, without soils to support plant growth. ■





CRISIS COUNSELLORS

Volcanic eruptions, oil spills and bacterial outbreaks all land in the laps of government science advisers, and put them to the test.

BY ALEXANDRA WITZE, LAUREN MORELLO & MARIAN TURNER

As a population biologist, John Beddington spent most of his career studying fisheries rather than worrying about volcanoes. But then came April 2010, when Beddington — the UK government's chief scientific adviser — found himself having to figure out not only how to pronounce Eyjafjallajökull, but also what to do about the eruption of the Icelandic volcano.

In the small hours of 14 April, the volcano had gone from its previous state — picturesquely spitting out lava — to violently spewing plumes of ash high into the atmosphere. Winds were blowing to the south and east, where the fine ash presented a threat to Europe's busy commercial airline routes. Suddenly scientists were scrambling not only to understand how much ash the volcano was generating, but also

how it was spreading through the atmosphere and how much of a risk it presented to aircraft. So Beddington got a call at his Cotswolds home summoning him to 10 Downing Street. "I sort of dusted off my brain and went into the meeting," says Beddington, who is now at the University of Oxford.

In the first week of the crisis, authorities progressively closed airspace where the volcanic ash was billowing. Ultimately more than 300 airports were shuttered across Europe, stranding some 8.5 million passengers and causing major economic losses to the airlines and businesses that depend on them. Each country made the decision about its own airspace, which put Beddington front and centre of helping UK officials figure out what to do.

When scientists enter government in the role of a scientific adviser or as the head of

a science agency, they need to be prepared for the unexpected. Some of their most crucial contributions come during crises, a theme that will be explored on 28–29 August at a global summit of science advisers in Auckland, New Zealand. On the eve of that meeting, *Nature* takes a look at how such officials performed during the Eyjafjallajökull eruption, as well as the 2010 oil spill in the Gulf of Mexico and a deadly disease outbreak in Europe the following year.

These cases show that science advisers have key roles in a crisis, especially in disseminating clear, reliable information to government leaders and the public. But at times, they struggle with the demands presented by disasters: rare events can take them by surprise, bureaucracy

The eruption of Eyjafjallajökull in Iceland stranded millions of people around the world.

ORSOLVA HAARBERG/NATURE PICTURE LIBRARY

can strangle their attempts to respond and they often cannot keep pace with the evolving situation. “We have to form a view about advice for the government,” Beddington says. “And we have to do that on a fairly quick time scale.”

FLIGHT RISK

After leaving his meeting with the prime minister, Beddington began to round up a panel of volcanology and meteorology experts to form an Eyjafjallajökull-focused Scientific Advisory Group for Emergencies (SAGE), the UK government’s main mechanism for gathering technical advice and passing it along to decision-makers in crises.

The SAGE concept was born in the wake of the 1990s spread of bovine spongiform encephalopathy, or ‘mad cow’ disease. Beddington was the first UK chief science adviser to gather a SAGE group together, during a 2009 influenza pandemic. Because of that experience, he says, when the Eyjafjallajökull crisis began, “I knew the sort of people I’d need”.

The SAGE volcanic ash group met for the first time on 21 April, after London’s Heathrow airport — the world’s busiest — had faced the cancellation of more than 97% of its flights for five days straight. The group included Sue Loughlin, a volcanologist at the British Geological Survey in Edinburgh, who did her PhD on Eyjafjallajökull and had served in Montserrat, in the West Indies, during a deadly eruption there in 1997. Loughlin and others supplied basic information about the volcano’s geological history and the pace of the ongoing eruption.

Yet the ash cloud, and the crisis, moved faster than the advisory group. Pressured by airlines

that wanted to resume flights, Europe’s transport ministers had on 19 April quickly brokered revised operational guidelines. Planes were in the air again even as SAGE began to meet.

After the initial eruption had quietened, the volcano continued to spew low levels of ash, and nobody knew whether the activity might pick up again. Three more times over the next two months, Beddington convened SAGE to assess technical details about the eruption and the likelihood of more to come. Among other things, the group explored whether a nearby volcano named Katla might also erupt, as it has in the past along with Eyjafjallajökull. The advisers passed that information to the Cabinet Office, which used it to develop scenarios for future volcanic-ash emergencies.

SAGE also pushed government departments to assess the risk of future, larger volcanic eruptions. In 2012, the Cabinet Office added Icelandic eruptions to Britain’s National Risk Register, the official list of possible events that could disrupt society. “I had not thought of it at all before then,” says Beddington. “It was very embarrassing.” The Cabinet Office is also working up a detailed scenario for how to respond in the face of an eruption that could spew sulphur and other toxic gases across Britain, as the Icelandic volcano Laki did for eight months in 1783–84.

David Alexander, a risk expert at University College London, says that the Eyjafjallajökull experience improved some aspects of disaster response in the country. The International Civil Aviation Organization, for example, has updated and clarified its guidelines on how much ash planes can fly through safely. And the UK Met Office has fine-tuned its atmospheric

models for predicting the spread of dry ash through the air.

But nearly all parts of the government took much too long to respond to the crisis, Alexander says. And he notes that even now, no coordinated plan exists to manage alternative transportation, such as the ferries, trains and taxis that became overloaded in April 2010. “There is still no adequate way for dealing with millions of stranded people,” Alexander says.

Before Beddington left office in 2013 — replaced by Mark Walport — he activated the SAGE mechanism once more, this time to provide advice about whether to evacuate the British nationals in Japan after the 2011 meltdown at the Fukushima nuclear power plant. SAGE modelled how radioactive material might spread and concluded that the risk of being exposed to radiation was relatively modest. In the end, the government provided iodine tablets as a precautionary measure but told its nationals they could stay put.

And just in case, Beddington also ran some tabletop exercises for a major space weather event that could blow out power grids as well as other events so alarming that he prefers not to even name them. “If any of these instances had happened,” he says, “we’d have been in position to pull a SAGE team together.”

DEEP TROUBLE

Even as the volcanic ash cloud was spreading over Europe, science officials in the United States were struggling with their own crisis, one of the biggest ecological disasters in the nation’s history. It all started just after 9 p.m. on 20 April 2010, when an engineer aboard

WHP/ALAMY



During the 2010 Gulf oil spill, President Barack Obama met with scientists Steven Chu (far left), Jane Lubchenco and John Holdren, and coast guard commandant, Thad Allen.

the BP Deepwater Horizon oil rig in the Gulf of Mexico noticed an odd vibration. Minutes later, the rig exploded, killing 11 men and beginning a months-long effort to stanch the flow of oil from a damaged well on the sea floor and avert an environmental catastrophe.

By the time the leak was first plugged in early August 2010, an estimated 4.9 million barrels had gushed into the gulf — surpassing all previous marine spills — and in so doing had put US President Barack Obama's vaunted

between 25,000 and 100,000 barrels per day. "The estimates that kept coming up formally from the agencies were to me just too low," says Kate Moran, an oceanographer then working as a senior policy analyst at the OSTP. That error created problems for BP when it tried to cover the gushing wellhead with an iron dome, because the cap was unable to contain the volume of escaping oil and gas.

It took until 27 May for the committee assembled by the administration to deter-

oceanographer," says Ian MacDonald from Florida State University in Tallahassee. It took another 22 days before NOAA acknowledged the presence of the plume — a delay that fed mistrust of the agency by outside scientists.

By contrast, McNutt and Chu drew praise for their actions during the crisis. While McNutt helped to determine the amount of oil leaking into the gulf, Chu worked to stop the flow. He arrived in Houston on 12 May, accompanied by distinguished scientists recruited with Holdren's help, including some from JASON, a storied panel that advises the government on issues such as defence and energy. They and others quickly began challenging BP for more and better data about the state of the well. Chu convinced the oil company to monitor the wellhead using γ -ray imaging as well as temperature and pressure gauges, which provided the first direct measurements of the still-flowing oil.

Finally, on 19 September — 150 days after the initial explosion and well blowout — the Coast Guard declared an end to the disaster after engineers inserted a cement plug to permanently seal the well. Four years later, disappointment still lingers among many scientists about the way in which the Obama administration handled Deepwater Horizon. Stopping the oil's flow "was a huge effort, and people worked heroically and tirelessly", MacDonald says. But, he argues, the scientists leading government agencies should have acted more quickly and provided better information about the extent and nature of the spill. "At many critical junctures in the process," he says, "the government was on the wrong side of history."

OUTBREAK

A year after the Deepwater Horizon spill, doctors in Hamburg, Germany, put out a call for help. On 19 May 2011, three children at a city hospital were battling haemolytic uraemic syndrome, a life-threatening condition caused by a severe gastrointestinal infection with the bacterium *Escherichia coli*. Worried that the outbreak could spread, Hamburg health officials contacted the Robert Koch Institute (RKI) in Berlin — Germany's federal agency for disease control and prevention.

When three RKI epidemiologists arrived in Hamburg the next day, it was clear that something unusual was going on. Several other cases, including some in adults, had popped up at hospitals around the city and reports soon came in from other regions. What followed was Europe's worst recorded outbreak of *E. coli* infection. By the time the outbreak was declared over, some two months later, more than 3,800 people had developed acute gastrointestinal infections. Of those, 845 had progressed to HUS, and 54 had died. The correct source of the infection — fenugreek seeds from Egypt — was not identified to the public until 5 July.

Reinhard Burger, president of the RKI, says that the outbreak "was a good example of how

"AT MANY CRITICAL JUNCTURES IN THE PROCESS, THE GOVERNMENT WAS ON THE WRONG SIDE OF HISTORY."

first-term science team to its toughest test, one for which it has received mixed marks.

Deepwater Horizon was a daunting disaster, with crude oil gushing from a reservoir of unknown size at a depth of 1,500 metres. But for more than a month after the blowout, the administration vastly underestimated the amount of oil flowing into the gulf — a mistake that hampered efforts to cap the leaking well and undermined public confidence in the president's response to the crisis. Key science officials, including presidential science adviser John Holdren, were slow to correct the erroneous estimates — even as academic scientists argued that the spill was orders of magnitude larger than BP or the government had publicly stated.

On paper, the crisis seemed tailor-made for the all-star group of scientists that Obama had assembled after he took office in 2009. Holdren, a physicist, directed the White House Office of Science and Technology Policy (OSTP), while Jane Lubchenco, a marine ecologist, helmed the National Oceanic and Atmospheric Administration (NOAA). Geophysicist Marcia McNutt had stepped in to lead the US Geological Survey (USGS). And the Energy Department boasted Nobel-prize-winning physicist Steven Chu as its leader.

Holdren says that the government sought help from outside scientists within hours of the explosion. But in many respects, Obama and his science team moved too tentatively. It was not until 19 May — almost a month after the blowout — that the administration assembled a group of scientists and engineers, headed by McNutt, to revise the controversial flow-rate calculations. The oil flow was extremely hard to estimate because there was no direct way to measure the well's output, and industry and research scientists disagreed about how best to do it.

NOAA had long maintained that the flow was about 5,000 barrels a day, but independent scientists examining satellite imagery of the growing oil slick and BP video of the under-sea well had argued that the actual output lay

mine that the flow rates determined by NOAA and BP were indeed too low. The group estimated that 12,000–19,000 barrels of oil were spilling into the gulf each day. Ultimately, the government would arrive at a much larger figure — 62,000 barrels a day immediately after the blowout, dwindling to 53,000 barrels a day in August, before a temporary seal stopped the flow.

"The most difficult part about it was trying to understand what we were getting from BP, and whether we really understood the possible sources of error," says McNutt, who arrived at BP's oil-spill operation centre on 6 May and remained for the duration of the crisis. "It actually took some time, and maybe too much time, to realize the magnitude of the problem," says Larry Mayer, an oceanographer at the University of New Hampshire in Durham. He argues, however, that the response quickly improved in mid-May, when Holdren and other officials met with scientists from some of the country's top oceanography programmes, seeking access to equipment such as remotely operated underwater vehicles and ways to track the oil as it spread in a plume below the surface.

That plume was a source of great grief for NOAA chief Lubchenco, whose agency was mandated by US law to assist the Coast Guard in tracking the oil's path through the Gulf of Mexico and monitoring its effects. In mid-May, academic researchers reported finding masses of microscopic oil droplets 1,000–1,400 metres below the ocean surface, spreading outward for tens, possibly hundreds, of kilometres from the leaking wellhead. Almost immediately, Lubchenco issued a statement calling reports of those findings "misleading, premature, and in some cases, inaccurate", drawing a wave of criticism from oceanographers who

felt that she was unduly dismissive of important evidence.

"That was truly a head-slapping moment for me as an

NATURE.COM
For more on
scientific advice to
governments, see
go.nature.com/c2stjq

rapidly a new threat can appear and develop and affect the population". The event also illustrated the problems that can emerge when scientists at multiple agencies — reporting to different levels of government — respond to a public-health emergency.

The RKI's most pressing task was to identify the strain of bacterium that was causing the infection and then where it came from. *E. coli* infections are typically caused by contaminated foods, and the epidemiologists began by interviewing patients about their recent diet. "We knew within the first two days or so that the usual suspects — fresh milk and raw-meat products — were not the problem," says Burger. The only thing all patients seemed to have in common was that they had eaten salads containing fresh tomatoes, cucumber and lettuces, in northern Germany. On 25 May, the RKI, which reports to the federal health ministry, issued a joint statement with the federal institute for risk assessment (BfR), saying that these vegetables were associated with increased risk of infection.

Science advisers at the RKI say that the various federal agencies responsible for disease control and food safety worked together smoothly to assess the risks and communicate them to the public. But the coordination between them and agencies that reported to state governments was not nearly as efficient. As cases emerged throughout Germany, the chain of orderly communication cracked. The worst blunder came on 26 May, when Hamburg health senator Cornelia Prüfer-Storcks announced that the Hamburg Institute for Hygiene and Environment had discovered enterohaemorrhagic *E. coli* (a pathogenic class of *E. coli*) on Spanish cucumbers. But she had not consulted the RKI before making that statement. And tests a few days later revealed that the bacteria on the suspect cucumbers did not belong to the same strain found in patients.

The ongoing uncertainty about the source of the bacteria led to trade restrictions and large economic losses, particularly in Spain, where farmers found themselves unable to export cucumbers and other suspect produce. The European Commission eventually issued a €227-million (US\$302-million) payout to farmers from several countries.

Burger says that the Hamburg statement "damaged confidence in public announcements". The problem was compounded by the public's increasing hunger for information while epidemiological investigations were failing to turn up more-specific leads. "I was facing the media every day, but sometimes we had nothing new to say," he says.

Looking back, health officials say that some technological improvements could help. For instance, Burger says that using genetic sequencing rather than the current culture-based diagnostic techniques might have helped physicians to recognize the outbreak more quickly.



Local authorities in Germany incorrectly identified Spanish cucumbers as the source of deadly *E. coli* during an outbreak in 2011. Reinhard Burger, who heads the nation's federal agency for disease control and prevention, and other health officials worked to clear up the confusion.

The response was also hampered by a complicated reporting chain: when physicians confirm a case of a disease, they report it to local health authorities, who forward it to the RKI. Before the outbreak, that process could take up to 16 days. A change to the law last year means that reports must now reach the RKI within 5 days. But the notifications can still trickle in by phone, fax or e-mail. The federal health ministry is therefore developing an electronic system to provide faster and simultaneous notification for local and federal authorities.

Such a system would not bypass the problem that the RKI can act only at the request of state agencies — a rule that some feel should be changed. "When there is the impression that an outbreak is affecting more than one state, the RKI should have the right to start investigations on its own," says Ulrich Frei, medical director of the Charité hospital in Berlin, which handled several cases. "Health is one of only a couple of topics in which the German states still have much authority, and they're not very willing to transfer this to the federal level," he says.

As with natural disasters and human-caused crises, the outbreak points out the benefits of high-level coordination during emergencies. Peter Gluckman, New Zealand's chief science adviser, grappled with this issue when an earthquake levelled much of Christchurch in 2011. In the aftermath, competing scientific experts debated future risks in the media in a way that spurred confusion among government officials and the public, says Gluckman. His office spent weeks trying to get scientists to provide clear information about existing threats and uncertainties. A crisis demands scientific coordination, he says. "Often it is made worse by inconsistent communication." ■ SEE EDITORIAL P.347

Alexandra Witze, who wrote about Eyjafjallajökull, reports for *Nature* from Boulder, Colorado, and is the co-author of *Island on Fire*, a book about an Icelandic volcanic eruption in 1783. **Lauren Morello**, an editor with *Nature* in Washington DC, wrote about the oil spill. **Marian Turner**, an editor with *Nature* in London, covered the *E. coli* outbreak.

COMMENT

ASTROBIOLOGY Grand tour of the search for life emphasizes human insignificance **p.368**



FICTION Sex and death drive science-inflected shorts by Margaret Atwood **p.370**

PALAEONTOLOGY Private collections can mean fossils are in good hands **p.371**

OBITUARY Peter Marler, interpreter of animal language, remembered **p.372**

JONAS BENDIKSEN/MAGNUM



Residents in the flood-prone district of Kurigram, Bangladesh, move a community mosque to safer ground.

Model human adaptation to climate change

We can no longer ignore feedbacks between global warming and how people respond, say **Paul I. Palmer** and **Matthew J. Smith**.

Current models of Earth's climate capture physical and biophysical processes. But the planet has entered a new state: humans are adapting to, as well as causing, environmental changes. This major feedback must be modelled. Projections of the future climate based on simple economic narratives¹ — from cuts in greenhouse-gas emissions to unmitigated growth — are unrealistic.

Faced with droughts and rising sea levels, people alter their behaviour. Even if global climate policy is effective, and nations deliver

on ambitious green-energy-production and sustainability targets, societies will be different in a warmer world. People will move to places that are richer in resources, or stay where they are and be pushed further into poverty. Population growth, urbanization, migration² and conflict³ will compound reactions to global temperature rises.

To understand how events might unfold and what kinds of responses will be most effective, Earth-system models need to capture human-climate dynamics. It will be an enormous challenge: we are only beginning

to understand how people respond to their environment. But omitting human behaviour is like designing a bridge without accounting for traffic.

DECISIONS, DECISIONS

There are two main scientific challenges to modelling socio-economic responses to climate change. The first is describing how humans make decisions. The second is describing the relationships between humans and the physical and biophysical components of the Earth system. ►

► Decision theory approximates how a person or group makes a choice on the basis of their values, aspirations, health, uncertainties and rationality. Values may vary according to, for example, age, wealth, education, sex, culture and religion. One group of people might see adopting nuclear power as a practical way to avoid greenhouse-gas emissions; another would find it unpalatable. Choices, such as whether to move or stay in a location as average temperatures rise, will be influenced by events such as droughts, fires or floods. Outcomes and pay-offs might be as great as life or death.

To understand the underlying patterns, we need to collect behavioural statistics on grand scales. How do people of different backgrounds respond to extreme weather, for example? Under severe drought, do people in sub-Saharan Africa behave differently from those in southern Australia? How do the decisions made by lower- and middle-income families differ?

ASSESSING INFLUENCE

Some statistics can be drawn from existing databases, such as health and education data compiled by the United Nations or national censuses. Measures of social and physical mobility can be extracted from social media and Global Positioning System (GPS) information from mobile phones without raising privacy issues. A 2008 study found that the distances travelled in six months by 100,000 mobile-phone users were not random, but were well described by a simple function⁴.

Historical studies can inform us about possible future scenarios. For instance, Peruvian civilizations migrated between coastal valleys and the Andean highlands between AD 640 and 1200. Historical climate records and mitochondrial DNA sequences extracted from the bones of people who lived there showed that coastal droughts and flash floods pushed the populations to move to the mountains, from where they returned when droughts afflicted those lands⁵. Other studies have reported a higher incidence of conflict since 1950 among communities experiencing changes in climate — cases rose by a few per cent for each standard-deviation increase in temperature or rainfall³.

More surveys and longitudinal studies are needed to find out what sorts of risks individuals say they will react to most strongly, and how they actually respond. For example, a study of migration intentions in rural Pakistan from 1991 to 2012 showed that extreme heat influenced people's choices but flooding did not⁶. The main reason was that heat waves do not attract as much financial compensation as flood relief.

We must also find out which measurements can act as reliable indicators for a broad spectrum of behaviours. For example, income can be a strong predictor for rates of

migration, conflict and reproduction, but alone, income is insufficient to predict any one of these behaviours.

And individuals do not always make rational or independent decisions. People often rely on simple rules to form judgements⁷. They might assume that others know better than them and thus follow social norms. Friends and relatives may hold sway, and news, online and social media are shared sources of information and influence.

Alternatively, individuals might act independently and rationally in the short term, yet collectively destroy a shared resource in the long term, as with climate change. People often have to make decisions with too little or

too much information. Taking out insurance against flooding, for instance, might have to be done without knowing the future likelihood of extreme rainfall.

It is therefore essential to understand how the individual decisions combine across networks to produce macroscopic behaviours. A challenge for modellers will be to find which scales of social structure (governments, parties, tribes, clubs or neighbourhoods) exert most influence, as well as accounting for regional differences.

A range of modelling techniques will be necessary. One is the agent-based approach. Thousands or millions of 'agents' — digital proxies for individuals — encoded with attributes such as age or wealth interact according to simple rules. Complex group responses, such as the dynamics of infectious-disease spread, can emerge. But outcomes depend on which attributes are assigned so modellers need to ensure that the chosen variables are the key determinants of a particular response in real life. Rather than be a response to the physical effects of heat, a war in a warming region might reflect the decline in local economic conditions because of crop failure or an influx of people from a neighbouring region.

Inference methods (such as approximate Bayesian computation) could be used to discover which individual attributes contribute most to group behaviour. Methods for modelling groups of individuals with similar characteristics can still generate emergent mass responses, revealing, for example, that the contrasting dispersal of animals on land and in oceans contributes to the observed contrasts in ecosystems⁸.

TAKE STOCK

Three main things are needed to model human responses to climate change: interdisciplinary research, appropriate computational and conceptual frameworks, and better data.

Collaborations between natural and social scientists should be facilitated by physical and virtual centres. A hub such as the Isaac Newton Institute for Mathematical Sciences in Cambridge, UK, where mathematicians come together to solve problems that require insights from many fields, could focus efforts for developing human–climate population-modelling techniques.

Journals and funding bodies will need to smile more on such collaborations. The next (sixth) report of the Intergovernmental Panel on Climate Change should justify the importance of improved understanding and modelling of societal responses to climate change by reviewing the evidence across different working-group reports.

Scientists will need to obtain new data sets and consolidate existing ones on decision-making processes associated with environmental change. With so many factors, new approaches will be needed alongside conventional methods to collect data across a spectrum of spatial and temporal scales. This will involve data collection through social media, surveys and GPS, and will require a degree of analytical integration that will challenge computational scientists.

A parallel issue is how to store and analyse large volumes of heterogeneous data. Much data relating to individuals (such as social-network structures and movement patterns) will come with sensitive commercial or privacy issues. Careful licensing arrangements between businesses, governments and academic institutions will be needed.

Ultimately, we must establish an international data-collection effort involving the public, private and voluntary sectors. Much as we take global stock of forests or biodiversity, we should regularly assess how people are being changed by the climate that they are changing. ■

Paul I. Palmer is professor of quantitative Earth observation at the University of Edinburgh, UK. **Matthew J. Smith** is an ecologist in the Computational Science Lab at Microsoft Research in Cambridge, UK. e-mail: pip@ed.ac.uk

1. Moss, R. H. *et al.* *Nature* **463**, 747–756 (2010).
2. *Foresight: Migration and Global Environmental Change Final Project Report* (Government Office for Science, 2011); available at <http://go.nature.com/somswg>.
3. Hsiang, S. M., Burke, M. & Miguel, E. *Science* <http://dx.doi.org/10.1126/science.1235367> (2013).
4. González, M. C., Hidalgo, C. A. & Barabási, A. L. *Nature* **453**, 779–782 (2008).
5. Fehren-Schmitz, L. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 9443–9448 (2014).
6. Mueller, V., Gray, C. & Kosec, K. *Nature Clim. Change* **4**, 182–185 (2014).
7. Kahneman, D. & Tversky, A. *Cogn. Psychol.* **3**, 430–454 (1972).
8. Harfoot, M. B. J. *et al.* *PLoS Biol.* **12**, e1001841 (2014).



An artist's impression of the Milky Way — one of hundreds of billions of galaxies in the Universe.

ASTROBIOLOGY

Cosmic prestige

Mario Livio welcomes a lucid description of attempts to evaluate how special humans are.

All the astronomical discoveries made since Nicolaus Copernicus demoted Earth from its position at the centre of the Universe have continued to erode humanity's perceived physical significance in the grand scheme of things. Consider this sequence of events: in 1920, US astronomer Harlow Shapley showed that the Solar System does not occupy the centre of the Milky Way, but is about two-thirds of the way out. Then, Edwin Hubble discovered that there are many other galaxies — a few hundred billion in the observable Universe, according to the latest observations (made, fittingly, by the Hubble Space Telescope). Next, it was found that even the stuff we are made of — ordinary baryonic matter — constitutes less than 5% of the Universe's energy budget. To top it all, speculative models based on cosmic inflation and string theory suggest that our entire Universe may be but one member of a 'multiverse', a huge ensemble of some 10^{500} universes.

On the planetary scale, there has also been an explosion of discoveries. Until 1992, there had been no confirmed discoveries of any planets outside the Solar System. However, observations since then (especially by the Kepler satellite) suggest that about 20% of all Sun-like stars in our Galaxy harbour approximately Earth-sized planets orbiting in the stars' 'habitable zones' — the regions of space that are neither too hot nor too cold, allowing liquid water to exist on a planet's solid surface. Given that liquid water is considered



The Copernicus Complex: The Quest for Our Cosmic (In) Significance
CALEB SCHARF
Allen Lane/Farrar, Straus and Giroux: 2014.

potentially a necessary ingredient for life, these statistics are (at the very least) promising for those who believe that there could be life elsewhere.

That Earth hosts life remains its last qualification for being special. How reasonable is it to think that we are alone in the vast expanses of space? And how significant is life on Earth on the Universal (or multi-versal) scale? These are the questions that astrobiologist Caleb Scharf addresses intelligently and comprehensively in his beautifully written *The Copernicus Complex*. The book offers a grand tour of important findings from astronomy to biology that are relevant to the cosmic and microscopic search for life.

What sets this book apart from those that simply describe the hunt for exoplanets is Scharf's emphasis on the significance, or lack thereof, of our own existence. For instance, the realization that the human body contains ten times as many microbial cells as human

NASA/ESA/Z. LEVAY (STSCI/AURA)

cells — as well as impressive advances in the understanding of the chemical origin of life — has forced us to rethink how we classify the ‘importance’ of life forms on Earth, and perhaps even to consider placing microbes at the top of the hierarchy, rather than at the bottom.

Scharf tackles in some detail the question of whether we can conclude anything about the expected frequency of extraterrestrial life from the known facts about the emergence and evolution of life on Earth, particularly two important clues. The first is that some

“Our place in the universe is special but not significant, unique but not exceptional.”

form of life arose very early in Earth’s history, within only a few hundred million years of the planet’s formation; the second, that the appearance of

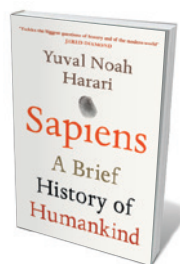
‘intelligent’ beings took a few billion years. After sketching the basics of Bayesian probability theory, Scharf describes the interesting results of astrophysicists David Spiegel and Edwin Turner. They have shown that in the absence (so far) of any evidence of life arising independently of our lineage, one cannot reach any conclusions about the rarity (or not) of life in the Universe. This highlights the importance of the search for that evidence.

Scharf ends his book with the reflection that life inhabits the border between order and chaos. For instance, the dynamics of the planetary orbits in our Solar System are so complicated that they may become unstable within a few billion years. Similarly, Earth’s climate and geophysics occupy that interface between order and disorder. From that, Scharf concludes that “our place in the universe is special but not significant, unique but not exceptional”. Note, however, that from the perspective of thermodynamics (entropy), life itself is an extremely ordered system.

I see two other important messages, with which I am sure Scharf would agree. One is that given the number of space telescopes that either are being built (such as the James Webb Space Telescope, to be launched in 2018) or have been proposed (including the Advanced Technology Large-Aperture Space Telescope), we may, for the first time in human history, be close to determining whether extraterrestrial life exists. Some optimistic estimates predict such a discovery in the next two decades. The second is that, notwithstanding our physical insignificance, the human mind is significant. Why? Because all the discoveries described in this book, from the subatomic realm to the multiverse, were made by us. ■

Mario Livio is an astrophysicist at the Space Telescope Science Institute in Baltimore, Maryland. His most recent book is *Brilliant Blunders*.
e-mail: mlivio@stsci.edu

Books in brief



Sapiens: A Brief History of Humankind

Yuval Noah Harari HARVILL SECKER (2014)

This newly translated Israeli best-seller by historian Yuval Noah Harari delivers a boldly synthesized account of *Homo sapiens*’ rise through the hominin ranks, by way of the cognitive, agricultural and scientific revolutions. Harari offers some original reframing of phenomena such as symbolic thinking. A leitmotif of cruelty in domains from war to livestock rearing also emerges — perhaps unsurprisingly in a species that, Harari argues, reached the top of the food chain by acting like the dictator of a banana republic. A view of our ascent as nasty, brutish, long — and endlessly fascinating.



A Prescription for Psychiatry: Why We Need a Whole New Approach to Mental Health and Wellbeing

Peter Kinderman PALGRAVE MACMILLAN (2014)

The furore over the 2013 edition of the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders* showed anew the rifts in psychiatry over diagnostic hair-splitting and medical interventions. Here, psychologist Peter Kinderman enters the fray. Arguing against biomedical reductionism, he offers a social and psychological model of mental illness, and calls for reforms such as multidisciplinary care, reduced use of pharmaceuticals, and big societal changes to promote mental well-being.



The New Moon: Water, Exploration, and Future Habitation

Arlin Crotts CAMBRIDGE UNIVERSITY PRESS (2014)

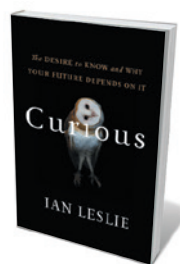
Is lunar exploration an old story? Pushing aside the political hurdles that impede a US return to the Moon, astrophysicist Arlin Crotts mines lunar research and its implications for human colonization in staggering, often deeply engaging, detail. Beginning with a scientific portrait of Earth’s satellite, he probes missions from the cold war and international space activity since then; delves into findings on lunar chemistry, the Moon’s far side and aspects such as outgassing and moonquakes; and explores in technical but accessible detail what we can glean from all this regarding a human presence on the Moon.



The Bee: A Natural History

Noah Wilson-Rich IVY (2014)

Well over 100 million years ago, flowering plants debuted on Earth — and early wasps began to co-evolve with them into the bee. As key pollinators and providers of honey and wax, bees have buzzed their way into human history. But the natural history of solitary, bumble, honey and stingless bees is as gripping as our lengthy alliance, as urban beekeeper Noah Wilson-Rich and contributors show in this charming compilation. They cover evolution, biology (including a unique proboscis made of two organs), behaviours (such as honey bee “quacking”), the causes of catastrophic die-offs, and more.



Curious: The Desire to Know and Why Your Future Depends on It

Ian Leslie BASIC BOOKS (2014)

Deep questing is essential to keeping your cognitive edge and social intelligence razor-sharp throughout life, posits Ian Leslie. He weaves ample science into his exploration; the brain’s caudate nucleus, for example, is associated with romantic love as well as the urge to learn, putting the latter “onto the same pathway as our most primal pleasures”. Staying curious into old age, he argues, demands that we imitate Benjamin Franklin: “sweat the small stuff while thinking big”, and never stop experimenting. [Barbara Kiser](#)



ILLUSTRATION BY JOE WILSON

on a widowed author of cult fantasy fiction, Constance, the first tale is 'Alphinland' — named after the river Alph in Samuel Taylor Coleridge's poem *Kubla Khan*, a symbol of human mortality that flows into a "sunless sea". Alphinland is the fictional realm that Constance has crafted, with the artist's perennial aim of cheating sex and death by creating alternate universes. 'Revenant' and 'Dark Lady' trace the final days of Constance's once-youthful poet-lover and the insubstantial muse who succeeded her. Constance tries to imprison them all fictionally in Alphinland, but cannot halt time in the real world.

From here the tales venture into stranger territory. In 'Lusus Naturae', the genetic lottery goes awry, creating an erudite woman whose congenital abnormality makes her resemble a vampire. The eponymous 'Stone Mattress' was originally cooked up by Atwood as a campfire tale on a trip in the Canadian Arctic. In it, we follow the ageing Verna on a tour boat in the Arctic. She has made a career of marrying elderly men with weak hearts, then hastening their deaths by giving them tacit permission to satisfy every forbidden desire. Confronted by a sexual predator from her past, she graduates to full-on murder using a 2-billion-year-old stromatolite, the fossilized remains of biofilms created by some of the very earliest forms of life.

Only once does Atwood turn to a larger stage, in 'Torching the Dusties' — my personal favourite. In it we meet Wilma, who lives in a comfortable retirement community. She is afflicted with Charles Bonnet syndrome, in which visual hallucinations of remarkable clarity — in this case, colourfully dressed little people who dance on windowsills — emerge as sight fades. Wilma is doted on by Tobias, a courtly fading romantic. They live an idyllic life, until news begins to filter in of angry mobs torching retirement communities around the world. The dusties, as the mobs refer to the elderly, are taking up too much time, space and money. When crowds gather at the gate, Wilma and Tobias make a run for it — to what end, we can only imagine.

Atwood is an author of remarkable gifts. Poet, storyteller and scientific acolyte, she ponders the inevitable end that awaits us all, and with great courage flushes her quarry out into the open. I feel about *Stone Mattress* the way I imagine Atwood feels about life: with the end in sight, I am ready to begin again. ■

Paul L. McEuen is the John A. Newman Professor of Physical Science at Cornell University in Ithaca, New York, and director of the Kavli Institute at Cornell for Nanoscale Science. His scientific thriller *Spiral* was named best debut novel of 2012 by *International Thriller Writers*. e-mail: pmceuen@gmail.com

FICTION

Transgressive treats

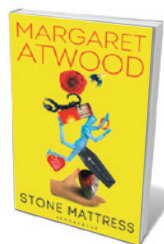
Paul L. McEuen relishes Margaret Atwood's acerbic tales of sex, hallucinations and death by stromatolite.

Sex and death are the pole stars of *Stone Mattress*, Margaret Atwood's fine new collection of nine dark, witty tales. The first three form a mini-trilogy; the remaining six constitute a smorgasbord of horror and crime inflected with science, not least characters with rare medical conditions, and a murderous use for one of Earth's oldest fossils.

Many of Atwood's most famous works, such as *The Handmaid's Tale* (1985, McClelland and Stewart) or the MaddAddam trilogy, look at individuals struggling under the yoke of a technologically fractured, dystopian society. *Stone Mattress* follows in this tradition, but the oppressor here is sex, along with its dark partner, death.

Sex is an innovation as old and

revolutionary as eukaryotic life itself. Each new generation is a risky shuffle of the genome, rolling the dice of variation in spectacular fashion. To complete the deal, the previous model needs to exit the stage — a strategy that is successful for the species, but wreaks havoc on the individual. So the trio of stories at the start centres on elderly ex-lovers for whom death lies directly ahead, and sex is a signpost in the rear-view mirror. Centring



Stone Mattress: Nine Tales
MARGARET ATWOOD
Bloomsbury/Nan
A. Talese: 2014

Correspondence

Uphold China's plan for cleaning water

We disagree with aspects of Tao Tao and Kunlun Xin's plan for sustainably producing China's drinking water (*Nature* **511**, 527–528; 2014).

In our view, the country should continue to improve its mains-water infrastructure: upgraded pipework is essential to its urbanization strategy (see X. Bai *et al.* *Nature* **509**, 158–160; 2014).

Projects are already under way for recycling water and cleaning water sources. In Beijing, for example, recycled water is used in industrial processes, for replenishing rivers and lakes, and for irrigation and domestic purposes. And Yuanping city in Shanxi province now has a purification plant for industrial wastewater.

We question the feasibility of installing household water purifiers, a core feature of the authors' plan. Each would cost around 1,500 renminbi (US\$244), which amounts to 645 billion renminbi for all the families in China — much more than is needed to upgrade pipes and water-treatment plants. Door-to-door recycling of filters could prove harder to manage than centralizing purification of household water in treatment plants. Also, filters in domestic purifiers could be a health risk if not changed frequently enough.

People's reluctance to switch from traditional ways or to use poor-quality water for tasks such as laundry could also undermine the success of the proposed plans. **Yanhong Tang** *Northeast Agricultural University, Harbin, China.*

Xin Miao *Harbin Institute of Technology, China.*
tangyanhong@aliyun.com

When less means more on dairy farms

Vanishing profit margins in dairy farming are intensifying pressure on the ubiquitous, genetically

elite, high-yielding Holstein milk cow. A radical change in strategy is needed — less-intensive agriculture and a reduction in consumer wastage (see go.nature.com/bwichl) could more than compensate for lost production.

The Holstein's remarkable milk production (up to 10,000 litres a year) is associated with poor condition, fertility and survival (see P. Dillon *et al.* *Livestock Sci.* **99**, 141–158; 2006), and has welfare implications. Its large disease burden demands routine hormone and antibiotic treatment, despite concerns about antimicrobial resistance. During a single lactation, one cow can eat more than its own body weight in cereals, much of which is potential human food and is grown using polluting artificial fertilizer.

An alternative would be to use cattle breeds that are genetically more resilient, less disease-prone and that have male calves suitable for beef production (females are used as milking-herd replacements). Cereal supplementation of feed may be minimized by taking advantage of the ruminant's ability to digest forage, green waste and fibrous by-products. Such a cow can still produce up to 8,000 litres of milk annually.

As the Pareto principle, or '80:20 rule', predicts, that would amount to around 80% of yield for just 20% of the environmental and welfare costs.

Mark C. Eisler, Michael R. F. Lee *University of Bristol, UK.*

Graeme B. Martin *University of Western Australia, Crawley, Australia.*

mark.eisler@bristol.ac.uk

Private collections of fossils are a plus

Paul Barrett and Martin Munt contend that private collections of fossil specimens hold back science because they are not readily accessible (*Nature* **512**, 28; 2014), but this need not be the case. The solution lies in closer

collaboration between private collectors and palaeontologists.

Private collectors provide a valuable service: many scientifically important specimens would never have been found, collected or prepared without their enthusiasm and dedication. This applies to every specimen of *Archaeopteryx* (the earliest bird) discovered so far. Public institutions, by contrast, often do not have the funds or staff to carry out essential excavations, or to acquire scientifically important specimens, which are not routinely donated.

Thanks to the cooperation of private collectors, such specimens can be well described and documented (using computed tomography scanning, for example). This is preferable to overlooking scientifically important data, even if access is subsequently limited.

We agree that specimens should be housed in conditions that allow verification of earlier observations. But this is not always the case for specimens held in public institutions: in our experience, access to material is sometimes denied, specimens may be lost or destroyed, or exhibited in such a way as to make detailed study difficult.

Oliver W. M. Rauhut, Adriana López-Arbarello, Gert Wörheide *Bavarian State Collection for Palaeontology and Geology; and Ludwig-Maximilians University of Munich, Germany.*
o.rauhut@lrz.uni-muenchen.de

Japan to learn from biomedical cases

The recent spate of high-profile retractions of biomedical papers by Japanese scientists is undermining the push by Japan's prime minister, Shinzo Abe, to strengthen medical research and innovation in the country (see go.nature.com/cznkrb). To stem this apparent proliferation of research misconduct, we suggest that Abe's government needs to

reform relations between Japan's pharmaceutical industry and its health and medical community.

Several cases have involved medical schools that receive unrestricted funding from big Japanese drug companies. These include the Jikei Heart Study, in which data manipulation was implicated in the clinical trial of a blood-pressure drug (see *Lancet* **382**, 843; 2013); the questionable procurement of patient data for a leukaemia drug by company employees (see J. McCurry *Lancet* **383**, 2111; 2014); and criticisms of data handling in a large government-supported study on Alzheimer's disease (see D. Normile *Science* **345**, 17; 2014).

There is also the infamous case of more than 100 papers by a Japanese anaesthesiologist that are now in the process of being retracted (see *Nature* **489**, 346–347; 2012).

These worrying examples may be symptomatic of a deeper malady. As well as being under pressure to publish, biomedical researchers in Japan are having to rely increasingly on large donations from the pharmaceutical industry as government funding shrinks.

The prime minister's 'Abenomics' strategy for better health care requires a global, fair and competitive yet collaborative environment and partnership among stakeholders. It is therefore imperative that investigations into misconduct allegations are openly and formally conducted and that the lessons learned are used to limit further cases.

Tetsuya Tanimoto, Masahiro Kami, Kenji Shibuya *University of Tokyo, Japan.*
tetanimot@yahoo.co.jp

CONTRIBUTIONS

Correspondence may be submitted to **correspondence@nature.com** after consulting the guidelines at **http://go.nature.com/cmchno**.

Peter Marler

(1928–2014)

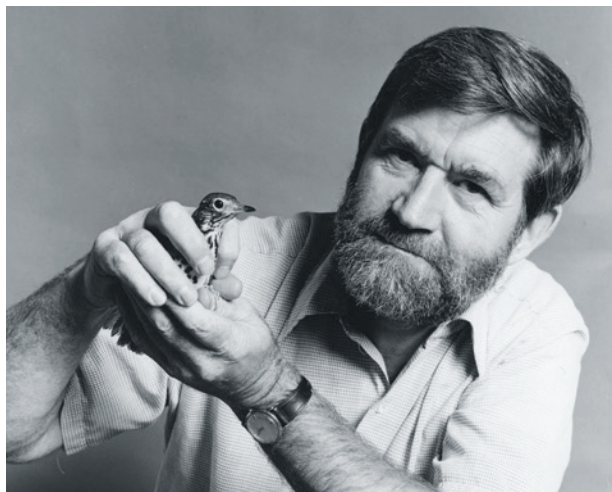
Pioneering interpreter of animal language.

White-crowned sparrows sing in distinct dialects in different parts of California. African vervet monkeys use various alarm calls to signal different dangers, such as snakes, mammalian predators or birds of prey. Peter Robert Marler chronicled these phenomena to establish ideas about how animals communicate. He hoped to find clues about the biology of human language. How does language acquisition blend innate knowledge and learning?

Marler, who died on 5 July, first became interested in animal sounds as a doctoral student in botany at University College London. Surveying potential nature reserves in Scotland, UK, he noticed that the song of the chaffinch *Fringilla coelebs* changed from valley to valley. This led to a second PhD, in zoology from the University of Cambridge, UK, in 1954, for which he described the complete vocal repertoire of this songbird, revealing that different chaffinch calls signalled different dangers. The observation was unprecedented.

Born in 1928 in Slough, near London, Marler left England in 1957 to join the faculty of the University of California, Berkeley. It was the heyday of ethology, the study of the behaviour of animals in their natural settings. Pioneered in Europe by Konrad Lorenz, Nikolaas Tinbergen and Karl von Frisch, ethology differed from work of comparative psychologists in the United States such as B. F. Skinner, who preferred to study animals — mostly rats and pigeons — in simplified laboratory settings. Whereas the psychologists sought universal rules of learning that applied to all animals, ethologists sought to understand animals' adaptations.

Marler, a field biologist, taught his students to imagine the challenges faced by wild animals, the resources and threats they encountered, the bonds they had to build, and the information they sought and conveyed. He challenged us to identify what signals they used in which contexts, and what responses they elicited. The song of a territorial chaffinch was a loud and ringing threat to rivals, but when the same male courted a female at close quarters, his voice morphed into a jumble of soft sounds reminiscent of the sweet talk of lovers. Notice these things, Marler taught us, but be careful with the inferences you draw.



I was enthralled when I first heard Marler lecture at Berkeley. Here was a young Charles Darwin building a rational edifice to understand how animals communicate. The goal was to discover how much of this process was instinctual, how much was learned and how had it all evolved.

Marler was superb at designing laboratory and field experiments. He used the latest techniques to record, analyse and play back sounds so that observers and instruments participated in 'conversations' with animals. After quantifying animals' responses, Marler would write his report, and there emerged his signature style. Taken together, he would say, "the observations suggested" some interpretation. It was his way of inviting readers and colleagues to join him in his exploration. No ringing truth, no strident dogma — it left him ample room to lead with ideas and observations while keeping the intellectual framework flexible.

Marler moved from Berkeley to Rockefeller University in New York in 1966, and in 1972 became the first director of the university's Center for Field Research in Ecology and Ethology in Millbrook, north of Manhattan. In this unique facility, he fostered the integration of field and laboratory work. In 1989, he moved his laboratory to the University of California, Davis, where he worked as a professor until retiring in 1994.

Marler encouraged his students to study any species that took their fancy: birds, flies, tarantulas, lizards, electric fish, bats, cetaceans, gerbils and wolves, among others. Students went to Kenya to study vervet monkeys, to Borneo to study orang-utans, and

to Japan to study macaques. Marler himself spent time studying colobus monkeys in Uganda and then, with primatologist Jane Goodall, the social behaviour of chimpanzees in Tanzania. He had hoped to glean insights on human language. He learned much about the signals that chimpanzees use, but felt that human language remained in its own class.

Marler astutely realized that if he shifted his focus from language to vocal learning, then birds had much to offer. Early on, he noticed that songbirds were picky about what sounds they imitated, and that choice often occurred during the first year of life. Moreover, the stages of vocal learning were set. This explained why

the dialects of his chaffinches disappeared when the young birds were reared away from wild-type models. Moreover, the what, when and how of vocal learning were directed, as if by an innate schoolteacher. Marler suggested that this interlocking of nature and nurture acted as an 'instinct to learn'.

This important insight was an alternative to the learning by trial and reward favoured by comparative psychologists. When it came to learning language, Marler felt that humans were more like his songbirds than like Skinner's rats — a conclusion that had repercussions. If language acquisition, so central to our manner of thinking, is ruled by biological predisposition, what does this tell us about the nature of human knowledge? Is it as idiosyncratic as the dialect of a chaffinch? At a time when much of the buzz in biology was molecular, Marler was one of the few philosophers of nature.

Peter was a gentleman. He adored his wife Judith and their three children. For many years, she ran the encampments where Peter and his students went to collect nesting songbirds, and she established home nurseries for hand-rearing the young. Peter and Judith were extraordinarily gracious hosts. They loved good food, good booze, good conversation and a good party. It was the good luck of Marler's students to be part of that world. ■

Fernando Nottebohm is professor of animal behaviour at Rockefeller University in New York, USA. He earned his PhD under Peter Marler at the University of California, Berkeley, in the 1960s.
e-mail: nottebo@mail.rockefeller.edu

COURTESY OF ROCKEFELLER ARCHIVE CENTER

GENOMICS

Hiding in plain sight

The latest releases from the ENCODE and modENCODE research consortia more than double the number of data sets on functional elements in the worm, fly and human genomes. [SEE ARTICLES P.393, P.400 & LETTERS P.445, P.449, P.453](#)

FELIX MUERDTER & ALEXANDER STARK

One of the major scientific achievements of our time has been the sequencing of the human genome and those of model organisms such as fruit flies and worms. These sequences encode species-specific information about protein-coding and non-coding genes and the regulatory information that determines when and where the genes are activated. However, even though this genomic information is present in the sequences, understanding it, or even just comprehensively identifying and annotating the different functional elements, is a major challenge. In an effort to identify all functional elements in the genomes of humans, *Drosophila melanogaster* flies and *Caenorhabditis elegans* worms, the Encyclopedia of DNA Elements (ENCODE) and the Model Organism ENCODE (modENCODE) research projects were launched^{1,2}. This issue of *Nature* contains five papers^{3–7} that summarize the latest data from these consortia. Together, the publications add more than 1,600 new data sets, bringing the total number of data sets from ENCODE and modENCODE to around 3,300 (Fig. 1).

The potential impact of such data is undeniable. More-complete genome annotations will form the basis for improved genetic studies in *D. melanogaster* and *C. elegans* — organisms that have already contributed most to our understanding of animal development and the molecular mechanisms involved. It is also increasingly clear that gene-regulatory elements are crucial for development and are frequently linked to disease; comprehensive identification of these elements should, for example, allow the interpretation of disease-associated mutations in non-coding genomic regions.

Two of the papers present data on RNA transcripts — Brown *et al.*³ (page 393) in *Drosophila* and Gerstein *et al.*⁴ (page 445) in all three species. Brown and colleagues' analysis of the *Drosophila* transcriptome, which they assessed in 29 tissues, 24 cell lines and 21 whole-animal samples that had been subjected to environmental perturbations, yielded more than 300,000 transcripts for 17,564 genes, of which 14,692 were protein-coding (different transcripts from the same

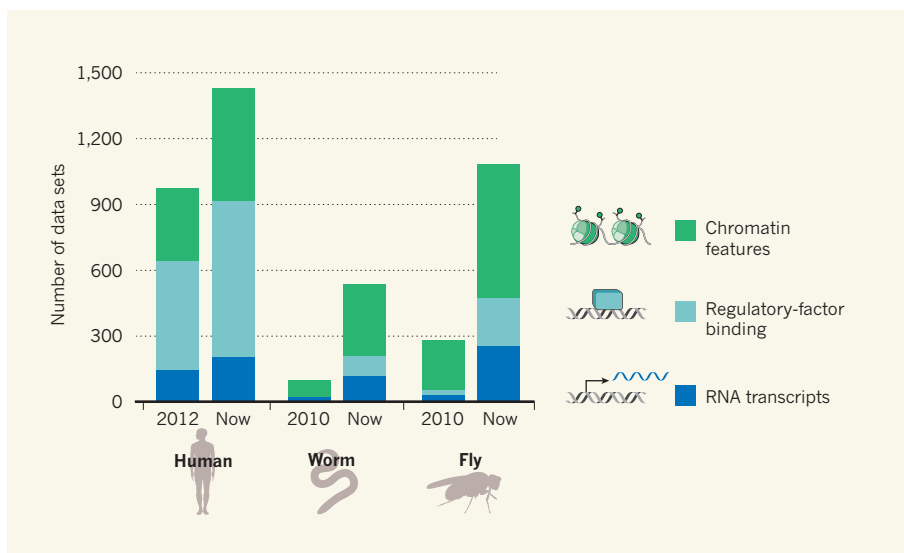


Figure 1 | The growth of ENCODE and modENCODE data sets. The ENCODE and modENCODE research consortia aim to identify all functional elements in the human genome and the genomes of the model organisms *Caenorhabditis elegans* (worm) and *Drosophila melanogaster* (fruit fly). The latest release^{3–7} from these projects focuses on three key data types: RNA-seq, which identifies RNA transcripts from cells or whole organisms; ChIP-seq for regulatory factors, which identifies locations in the genome that are bound by these proteins; and sequencing-based assays to profile various features of chromatin (the complex of DNA and histone proteins). The graph shows the total number of data sets now available for these data types, compared with previous releases^{19–21} (note that the numbers for the previous worm and fly releases do not include some microarray-based data sets).

gene are referred to as transcript isoforms). Of these genes, 57 (5,259 transcripts) were expressed only during perturbations and would thus probably escape identification under standard laboratory conditions. The analysis also identified many new candidate long non-coding RNAs, including ones that overlap with previously defined mutations that have been associated with developmental defects. Another intriguing finding was a small number of mostly neuronal genes that give rise to half of all detected transcript isoforms, reminiscent of the many transcripts known to be generated from the neuronal gene *Dscam*⁸. These data show that sampling selected tissues under non-standard conditions allows new genes and transcript isoforms to be identified even in well-studied organisms.

Regulatory elements are more difficult to identify than transcripts. They are typically predicted on the basis of characteristic features of chromatin (the complex of histone proteins and DNA in the cell nucleus) and by

studying regulatory-protein binding to DNA⁹ — refining such predictions is a key aim of both the ENCODE and modENCODE projects. Among the latest releases, Araya *et al.*⁵ (page 400) report the genome-wide binding profiles for 92 regulatory proteins, including transcription factors, RNA-polymerase subunits and chromatin-associated factors, in whole embryos and larvae from different developmental stages in *C. elegans*. Although this approach may provide information on regulatory changes during development, it is limited by a lack of cellular resolution¹⁰: transcription factors typically associate with cell-type-specific partner proteins to bind to different sites and regulate distinct genes in different cell types. Therefore, targets that are bound in only a few cells could be missed in whole-organism studies, and those that are found may constitute a superposition of binding sites from different cells. The authors partly deconvoluted these by determining the expression patterns for 180 genes, including

13 of the transcription factors profiled, in the early embryo at single-cell resolution.

Araya and colleagues' data also include binding profiles for predicted transcription factors that are otherwise uncharacterized. This will allow hypotheses to be generated about the proteins' possible functions, particularly, for example, if the binding sites are enriched near certain types of gene^{11,12}.

A key feature of this rollout of ENCODE and modENCODE data are comparisons across the three species studied. Complementing Araya and colleagues' data in worms, Boyle *et al.*⁶ (page 453) present almost 500 new genome-wide binding maps for transcription-regulatory factors in human cell lines, *Drosophila* and *C. elegans*. They found that about half of the binding events in each species occur at high-occupancy target (HOT) regions^{13,14}, where binding is heavily clustered. Although the function of these regions has not been assessed, our work in *Drosophila*¹⁵ suggests that many are active enhancers, which trigger gene transcription. However, because factors can bind DNA without functional consequences, especially at HOT regions, the contribution of each of the bound factors to enhancer activity remains unclear.

Apart from the existence of HOT regions, Boyle and colleagues' data reveal only a few commonalities between the species. But this is not unexpected — regulatory connections and target genes for individual transcription factors vary substantially between different cell types in a single species, so it is not surprising that there is little overlap in data derived from samples as disparate as human cell lines and whole fly and worm embryos. Thus, although the data sets may be valuable in each of the species, their usefulness for studying the evolution of gene regulation in cross-species comparisons is questionable, because such studies should compare homologous cell types that have shared developmental and functional properties.

Ho and colleagues' comparisons⁷ (page 449) focused on chromatin features that characterize regulatory genomic elements, such as DNA accessibility and certain modifications to histone proteins. In 800 new chromatin data sets, they identified several features common to the three species, including shared histone-modification patterns around genes and regulatory regions. Gerstein *et al.* integrated this information with transcription data to present a 'universal model' for predicting gene expression. As the authors point out, these commonalities are not surprising⁷ and are in agreement with the modifications' known distributions in each of the three species and in yeast. Instead, Ho and colleagues focused on the observed differences, which predominantly concern chromatin regions that are repressive (gene transcription from such regions is suppressed).

These five papers represent a substantial addition to the public ENCODE and

modENCODE resources. We expect the transcriptome data sets to have a direct influence on gene annotations in all three species, which should affect the work of many researchers immediately^{16,17}. It is arguably more difficult for scientists to easily access the data on chromatin features and regulatory-factor binding sites, and the regulatory-element predictions. This needs integration with the community portals^{16,17} and intuitive interfaces that allow data visualization and flexible analyses, which are being developed by the UCSC Genome Browser project and Ensembl, the two consortia, and others (such as i-cisTarget¹¹ or GREAT¹²). The success of the ENCODE and modENCODE resources depends on such interfaces being integrated into workflows throughout the research community.

Furthermore, although they are extremely data-rich, the papers expose how data sets that are created to catalogue all functional elements under standardized conditions are not sufficient for understanding the regulation of transcription, chromatin biology and enhancer function, nor the evolution of these mechanisms. Addressing such questions typically requires more-diverse set-ups and experiments, often specifically adjusted for each question. In addition, the identification of regulatory elements remains limited¹⁰ by the lack of cell-type specificity and the fact that chromatin features and regulatory-factor binding are imperfect predictors of regulatory-element function⁹. The papers do not reveal how many of these elements might be functional, and independent estimates span a broad range^{9,18}. However, the new data, in conjunction with the work of many other groups,

will undoubtedly aid future research into the identification, functional characterization and understanding of genes, regulatory elements and animal genomes more generally. ■

Felix Muerdter and Alexander Stark are at the Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), 1030 Vienna, Austria.

e-mail: stark@starklab.org

1. Celniker, S. E. *et al.* *Nature* **459**, 927–930 (2009).
2. The ENCODE Project Consortium. *Science* **306**, 636–640 (2004).
3. Brown, J. B. *et al.* *Nature* **512**, 393–399 (2014).
4. Gerstein, M. B. *et al.* *Nature* **512**, 445–448 (2014).
5. Araya, C. L. *et al.* *Nature* **512**, 400–405 (2014).
6. Boyle, A. P. *et al.* *Nature* **512**, 453–456 (2014).
7. Ho, J. W. K. *et al.* *Nature* **512**, 449–452 (2014).
8. Schmucker, D. *et al.* *Cell* **101**, 671–684 (2000).
9. Shlyueva, D., Stampfel, G. & Stark, A. *Nature Rev. Genet.* **15**, 272–286 (2014).
10. Furlong, E. E. M. *Nature* **471**, 458–459 (2011).
11. Herrmann, C., Van de Sande, B., Potier, D. & Aerts, S. *Nucleic Acids Res.* **40**, e114 (2012).
12. McLean, C. Y. *et al.* *Nature Biotechnol.* **28**, 495–501 (2010).
13. Moorman, C. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
14. The modENCODE Consortium *et al.* *Science* **330**, 1787–1797 (2010).
15. Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J. & Stark, A. *Genes Dev.* **26**, 908–913 (2012).
16. Yook, K. *et al.* *Nucleic Acids Res.* **40**, D735–D741 (2012).
17. St. Pierre, S. E., Ponting, L., Stefancsik, R., McQuilton, P. & the FlyBase Consortium. *Nucleic Acids Res.* **42**, D780–D788 (2014).
18. Kwasniewski, J. C., Fiore, C., Chaudhari, H. G. & Cohen, B. A. *Genome Res.* <http://dx.doi.org/10.1101/gr.173518.114> (2014).
19. The ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).
20. Gerstein, M. B. *et al.* *Science* **330**, 1775–1787 (2010).
21. The modENCODE Consortium *et al.* *Science* **330**, 1787–1797 (2010).

ASTROPHYSICS

Supernova seen through γ -ray eyes

Observations of γ -ray photons from a type Ia supernova indicate that stellar explosions of this kind get their energy from sudden thermonuclear fusion in the progenitor star. SEE LETTER P.406

ROBERT P. KIRSHNER

On page 406 of this issue, Churazov *et al.*¹ report a great discovery — not because it is a surprise, but precisely because it is not. The researchers have detected γ -ray emission lines from the type Ia supernova 2014J in the nearby galaxy M82 using the European Space Agency's INTEGRAL spacecraft. For decades, astronomers have been working out the physical picture for this type of exploding star on the basis of

the optical light it emits. The authors' study confirms directly the most fundamental idea in that picture by observing a supernova in the γ -ray range of the electromagnetic spectrum. The γ -rays they observed in the months after the supernova explosion were produced, as expected, by the radioactive decay of isotopes fused in a thermonuclear flame that destroyed a compact star.

Astronomers react quickly to explosive events. On the evening of 21 January 2014, while supervising an undergraduate astronomy



Figure 1 | Supernova SN 2014J in nearby galaxy M82. This image is a composite of optical images taken with the Hubble Space Telescope. Churazov *et al.*¹ have used the INTEGRAL spacecraft to observe γ -rays from SN 2014J.

laboratory at University College London, Steve Fossey and his students noticed an extra star in M82, later named supernova (SN) 2014J (Fig. 1). The star was detected by accident, but its discovery set off a flurry of observations to confirm that it was a supernova², determine its type and gather the most telling bits of data. Objects that are near the Sun have the most to teach us, and SN 2014J is the nearest type Ia supernova explosion since 1972.

There are two main kinds of supernova: those that get their energy from the gravitational collapse of their progenitor stars, and those whose energy comes from catastrophic thermonuclear fusion in the stars — the type Ia class to which SN 2014J belongs. These explosions take about 3 weeks to reach their peak of 4 billion times the Sun's light output. After a more-or-less symmetrical peak, they exhibit a long exponential decay. Their bright peak means that they can be seen halfway across the Universe. They have a narrow range in brightness, once corrections to their optical emission are applied, so they can be used as 'standard candles' to gauge cosmic distances with good precision.

That important application led to the 2011 Nobel Prize in Physics for the discovery that the expansion of the Universe is accelerating³. However, the fundamental idea that these supernovae result from the sudden thermonuclear fusion of carbon in a white-dwarf star has not had a direct test — until now. Although it is permissible to use objects that we do not fully understand to learn about the cosmos, it is much better to understand them. Churazov and colleagues have gone to the heart of the matter by directly

measuring the results of the nuclear fusion that destroyed the progenitor star of SN 2014J and made it glow.

Towards the end of its life, a star with the mass of the Sun will shed its outer layers as its core shrinks down to become a white dwarf. A white-dwarf star with the mass of the Sun is the same size as Earth. White dwarfs are durable if left alone. With no nuclear energy input and a small emitting surface, single white dwarfs just cool off slowly over time, gradually fading. But they are brittle — if a white dwarf acquires mass from a stellar companion, or smashes into another white dwarf, that extra weight can compress the carbon in the star's core until this element undergoes nuclear fusion. Fusion happens in a flash: a thermonuclear flame rips through the white dwarf, fusing carbon into heavier elements with a sudden release of energy that tears the star apart. Fusion stops yielding energy at the element that has the most tightly bound nucleus — in the case of a white dwarf, nickel-56 (^{56}Ni).

An exploding white dwarf should be full of ^{56}Ni nuclei — the pressure produced by the nuclear flame shreds the star and blasts these nuclei out at thousands of kilometres per second. In the first days after the explosion, the rise to the optical peak is powered by the radioactive decay of ^{56}Ni , which has a half-life of 6.1 days and decays into cobalt-56 (^{56}Co), which has a 77-day half-life. This isotope then decays into stable iron-56 (^{56}Fe ; ref. 4), helping to power the emission from the supernova. Type Ia supernovae that exploded 5 billion years ago are the cosmic sources of the iron in the Sun, Earth and your blood.

This is the conventional wisdom for

this class of supernova^{5,6}. Answers to my introductory astronomy exams get marked 'correct' if you write these things, but it is always better to look at the book of nature than at any textbook. The direct test of this canonical model would be to detect γ -rays emitted from ^{56}Co as it decays to ^{56}Fe in the weeks after the explosion. At first, most of the γ -rays are absorbed in the rapidly expanding material of the shredded white dwarf. This energy deposit accounts for the continuing luminosity of a type Ia supernova. To match the observed energy output of a type Ia supernova requires about 0.4–0.6 solar masses of ^{56}Ni in the expanding debris. Eventually, as the star's material thins out, some of the γ -rays begin to escape from the explosion without depositing their energy, so if you have the right kind of detector, you have a chance to observe them⁷.

Churazov and colleagues did not miss their chance. They observed SN 2014J between days 50 and 100 after the explosion using the INTEGRAL satellite and measured the γ -ray emission. As they show in their Figure 3 (see page 407 of the paper¹), there is a new spot of high-energy emission at the location of SN 2014J that was not present in 2013 (before the explosion occurred). Moreover, the γ -ray energy spectrum of the photons derived from that spot is a good match to a model spectrum of the γ -rays that escape from the exploding star (see Fig. 1 of the paper¹). The observed amount of γ -ray emission corresponds to about 0.6 solar masses of ^{56}Ni at the source, right in the ballpark of white-dwarf explosion models. The data are noisy because this measurement is at the limit of present technology, so it is not wise to trust too many details, but the basic results are clear: type Ia supernovae really are thermonuclear events.

Upsetting the conventional wisdom is always a joy in science. You can get prizes for that. But there is also a deep pleasure in showing decisive evidence on an important physical idea that has been used without proof for decades. The INTEGRAL data establish the thermonuclear explosion mechanism for type Ia supernovae. It is a wonderful result. ■

Robert P. Kirshner is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: rkirshner@cfa.harvard.edu

1. Churazov, E. *et al.* *Nature* **512**, 406–408 (2014).
2. Fossey, J., Cooke, B., Pollack, G., Wilde, M. & Wright, T. *Cent. Bur. Electron. Telegrams* **3792**, 1 (2014).
3. Kirshner, R. P. *The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos* (Princeton Univ. Press, 2004).
4. Arnett, W. D. *Astrophys. J.* **253**, 785–797 (1982).
5. Nomoto, K., Thielemann, F.-K. & Yokoi, K. *Astrophys. J.* **286**, 644–658 (1984).
6. Woosley, S. E. & Weaver, T. A. *Ann. Rev. Astron. Astrophys.* **24**, 205–253 (1986).
7. The, L.-S. & Burrows, A. *Astrophys. J.* **786**, 141 (2014).

IMMUNOLOGY

Mammalian watchdog targets bacteria

The aryl hydrocarbon receptor elicits protection against toxic environmental molecules. New data show that the receptor also supports the immune system by recognizing bacterially encoded virulence factors. [SEE ARTICLE P.387](#)

PARAG KUNDU & SVEN PETTERSSON

All animals, including humans, are exposed daily to a variety of chemicals in the air, water and food. Some of these carry valuable information about the host's environment, such as the presence of food, predators, members of the opposite sex or, in the case of hyenas, members of the clan to which they belong¹. But others are toxic and must be eliminated. Among several mechanisms for detecting and responding to these environmental cues is the aryl hydrocarbon receptor (AhR) protein, which can facilitate the biotransformation and elimination of toxic compounds encountered in the environment. On page 387 of this issue, Moura-Alves *et al.*² report that bacterial compounds known as phenazines also act as potential AhR ligands, and that recognition of these virulence factors by AhR contributes to host defence against invading microbial pathogens.

AhR is widely expressed in the mammalian body and is bound by a broad range of ligands that are mostly aromatic and hydrophobic compounds of endogenous or synthetic origin³. The unbound receptor is retained in an inactive form in the cellular cytoplasm but moves to the nucleus following ligand binding. Once in the nucleus, AhR has several functions, including marking sex-steroid receptors for destruction (by ubiquitination)⁴ and inducing the transcription of a battery of target genes involved in the regulation of cellular stress and metabolism⁵.

AhR has also been implicated in cross-talk with the immune system, particularly in promoting the differentiation of Th17 cells⁴. This suggests that the receptor may have a broad range of functions in addition to clearing unwanted chemical substances. Because the neutralization of microbial infections is one of the key functions of the mammalian immune system, Moura-Alves and colleagues used a molecular-modelling approach to test whether AhR senses ligands of bacterial origin. They found that pigmented virulence factors from pulmonary pathogens such as *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis* can bind to the ligand-binding domain of the receptor. They also provide evidence that this new class of ligand,

namely phenazines from *P. aeruginosa* and phthiocol from *M. tuberculosis*, activate AhR in a dose-dependent manner, leading to the elimination of these virulence factors, possibly through an AhR-controlled metabolic circuit (Fig. 1).

Phenazines and phthiocol are versatile secondary metabolites synthesized by bacteria and are known to influence bacterial interactions with their hosts. Most of these compounds have antibiotic properties and play a key part in regulating cellular redox states and the generation of reactive oxygen species, thus enhancing the virulence of their manufacturer⁶. In plants, phenazines influence growth by eliciting 'induced systemic resistance' and protection against plant pathogens^{6,7}. Thus, it seems that phenazine-producing bacteria display an interesting species-specific dichotomy, acting as symbionts in plants and as pathogens in animals. Although phenazine-producing bacteria are often soil or plant-dwellers, they are also found in the normal human microbiota⁸. For instance, *Nocardia* species are a part of our oral microbiota and reside in our healthy oral cavity, and *Methanococcina maei* is a component of our gut microbiome⁶.

The idea that microbial metabolites activate AhR is not new, as it has long been known that indoles, a group of AhR ligands, are generated by bacterial metabolism of the amino acid tryptophan⁹. Lactobacilli, found among our gut microbes, produce indole-3-aldehyde as a tryptophan metabolite, and this seems to act as an AhR ligand, promoting host resistance to fungal pathogens¹⁰. This suggests that phenazine-dependent activation of AhR may execute other functions in host responses to bacteria in addition to the clearance of virulence factors.

To validate their findings *in vivo*, Moura-Alves *et al.* studied mice lacking the gene encoding AhR, and found that infection with phenazine-producing *P. aeruginosa* induced more-aggressive disease characteristics and increased bacterial load compared to mice with AhR. They also identified two classes of cell — myeloid and parenchymal cells — as the major contributors to this AhR-mediated host defence. An additional twist to the story comes from their finding that phthiocol sensing by AhR, especially by myeloid cells, increases resistance to *M. tuberculosis* infection and prevents systemic dissemination of the bacterium in mice.

These findings establish a direct dialogue between AhR functions and invading pathogenic microbes, thereby consolidating the concept that AhR is an integral part of mammalian immunity. This new function of AhR is somewhat surprising, given previous demonstrations that AhR activation impairs immune responses to a variety of pathogens including the influenza virus¹¹ and herpesviruses¹². Moreover, because our normal, non-pathogenic microbiome contains phenazine-producing bacteria, there must exist a form of tolerance that allows maintenance of these populations

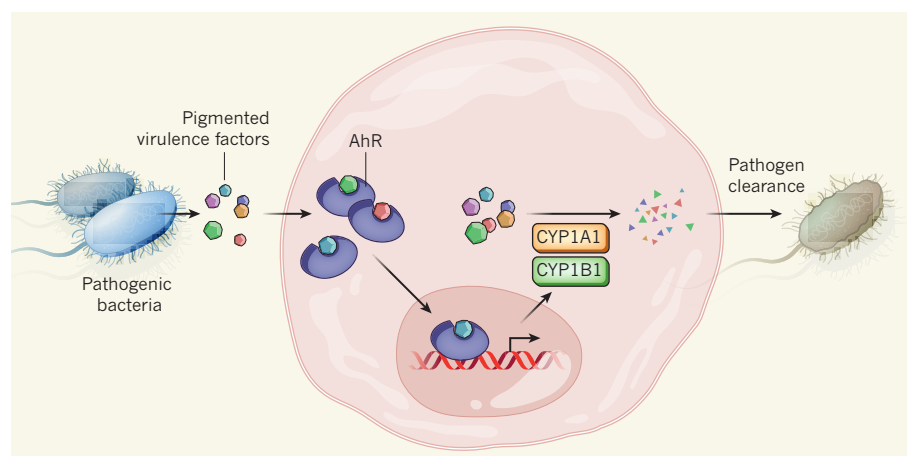


Figure 1 | AhR senses bacterial virulence factors and regulates host defence. Moura-Alves *et al.*² report that the mammalian aryl hydrocarbon receptor (AhR) senses pigmented bacterial virulence factors, including phenazines produced by *Pseudomonas aeruginosa* and phthiocol from *Mycobacterium tuberculosis*. Binding of these bacterial metabolites to AhR induces the receptor's movement to the nucleus, where it activates the transcription of genes for toxin-metabolizing enzymes such as CYP1A1 and CYP1B1. The authors suggest that AhR-induced increased expression of these enzymes eventually leads to the degradation of the virulence factors and subsequent clearance of the pathogens through host-defence mechanisms.

in specific sites, possibly mediated by an AhR-dependent mechanism. Indeed, a recent report portrays a 'disease-tolerance defence pathway' controlled by AhR, in which AhR-dependent tolerance against lipopolysaccharide, a component of the bacterial cell wall, imparts protection against pathogenic invasion¹⁵. Collectively, these findings provide evidence for AhR's role in mammalian host defence against phenazine-producing bacterial infections and unfold an exciting chapter in our understanding of AhR functions.

The diverse collection of AhR ligands, including hazardous chemical substances, metabolites from tryptophan, dietary ligands in fruits and cruciferous vegetables, and phenazines, suggest that this elusive 'Scarlet Pimpernel'-like receptor harbours a complex

and diverse repertoire of functions that remain to be discovered. Moura-Alves and colleagues' findings spark the fascinating idea of an evolutionarily developed AhR-microbiome connection, through which microbial communities can modulate host functions to reinstate the 'survival of the fittest'. ■

Parag Kundu and Sven Pettersson are in the Lee Kong Chian School of Medicine and the Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551. S.P. is also at the National Cancer Center, Singapore General Hospital, Singapore and in the Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden. e-mail: sven.pettersson@ki.se

1. Theis, K. R. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 19832–19837 (2013).
2. Moura-Alves, P. *et al.* *Nature* **512**, 387–392 (2014).
3. Denison, M. S. & Nagy, S. R. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334 (2003).
4. Ohtake, F. *et al.* *Nature* **446**, 562–566 (2007).
5. Stockinger, B., Di Meglio, P., Gialitakis, M. & Duarte, J. H. *Annu. Rev. Immunol.* **32**, 403–432 (2014).
6. Pierson, L. S. III & Pierson, E. A. *Appl. Microbiol. Biotechnol.* **86**, 1659–1670 (2010).
7. De Vleeschauwer, D., Cornelis, P. & Hofte, M. *Mol. Plant Microbe Interact.* **19**, 1406–1419 (2006).
8. Mavrodí, D. V. *et al.* *Appl. Environ. Microbiol.* **76**, 866–879 (2010).
9. Smith, T. J. *Exp. Med.* **2**, 543–547 (1897).
10. Zelante, T. *et al.* *Immunity* **39**, 372–385 (2013).
11. Jin, G. B., Moore, A. J., Head, J. L., Neumiller, J. J. & Lawrence, B. P. *Toxicol. Sci.* **116**, 514–522 (2010).
12. Inoue, H. *et al.* *J. Immunol.* **188**, 4654–4662 (2012).
13. Bessede, A. *et al.* *Nature* **511**, 184–190 (2014).

This article was published online on 13 August 2014.

NEUTRINO PHYSICS

What makes the Sun shine

Neutrinos produced in the nuclear reaction that triggers solar-energy generation have been detected. This milestone in the search for solar neutrinos required a deep underground detector of exceptional sensitivity. [SEE ARTICLE P.383](#)

WICK HAXTON

A remarkable detector of solar neutrinos called Borexino has operated for the past seven years in Italy's Gran Sasso Laboratory, shielded by more than a kilometre of rock from the cosmic rays that bombard Earth's surface. A prolonged effort has reduced background signals from radioactive elements present in the detector that would otherwise obscure the neutrino signal. On page 383 of this issue, the Borexino Collaboration¹ reports how this background suppression has enabled direct detection of the low-energy neutrinos produced in the nuclear reaction that initiates solar-energy generation.

Attempts to identify the source of solar energy have a long history². In the nineteenth century, gravitational contraction was the only known mechanism that could plausibly account for the Sun's luminosity and extended lifetime. Because this limited the Sun's age to about 30 million years, physicist William Thomson (Lord Kelvin) concluded that Charles Darwin's geological estimate of 300 million years for the age of the Earth was incorrect.

The beginning of the twentieth century marked the discoveries of the equivalence of mass and energy, embodied in Albert Einstein's famous equation $E = mc^2$, and of radioactive decays in which a nucleus changes its charge through weak interactions. Following Francis

Aston's measurements in 1920 of the mass difference between four protons and a helium-4 (⁴He) nucleus, Arthur Eddington proposed that the source of solar energy is the fusion of

four protons to form ⁴He. Later, responding to criticism that the Sun is not sufficiently hot to sustain nuclear fusion, Eddington invited his critic to "go and find a hotter place".

This dispute was resolved by George Gamow, who showed that quantum tunnelling would allow two solar protons to approach one another within the range required for nuclear fusion to occur. The detailed reactions leading to the synthesis of ⁴He were then deduced³: the proton-proton (pp) chain (Fig. 1) in the case of small, slowly evolving stars such as the Sun, and the carbon-nitrogen (CN) cycle in more-massive, rapidly evolving stars. Steady nuclear-energy release in the solar core keeps the temperatures high, ionizing hydrogen (H) and ⁴He and producing a plasma in which the electrons act as a gas. The Sun burns in a

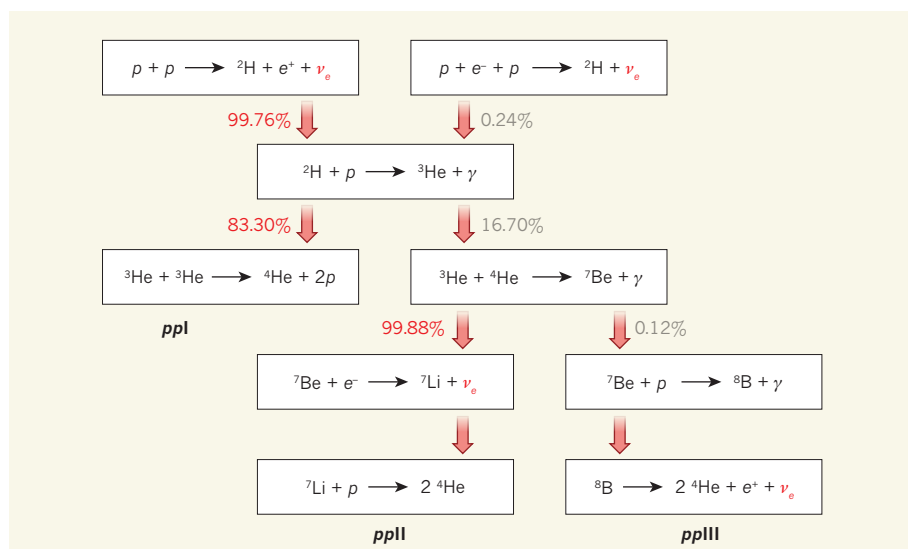


Figure 1 | The pp chain. The proton-proton (pp) chain of reactions is the dominant mechanism by which four protons are fused in the Sun to produce ⁴He. Its three cycles, labelled ppI, ppII and ppIII, are each accompanied by a distinctive electron neutrino (ν_e). In cases in which two reactions compete, the relative percentages of the two branches are indicated (taken from the standard solar model¹²). p, Proton; ²H, deuterium; e^+ , positron; e^- , electron; γ , γ -ray; ³He, helium-3 and -4; ⁷Be, beryllium-7; ⁷Li, lithium-7; ⁸B, boron-8. The Borexino Collaboration¹ reports the first measurement of the neutrinos associated with the $p+p$ reaction, and earlier measured the neutrinos from the ⁷Be and $p+e^-+p$ reactions.

long-lasting hydrostatic equilibrium, with the effects of gravity counterbalanced by the electron-gas pressure. The rate of energy generation must compensate for the radiative losses from the Sun's surface. The Sun has maintained this balance for 4.6 billion years by consuming about half the hydrogen fuel in its core.

This description of the Sun also applies to about 90% of the stars in the Milky Way. How can our understanding of energy generation in such 'main sequence' stars be tested? Conventional observations are limited to the Sun's surface: they do not directly probe the deep interior, where the temperatures of 10^7 kelvins necessary for fusion are found. Yet just as the Sun's surface shines in photons, its core shines in neutrinos. Each synthesized ^4He nucleus requires the conversion of two protons to two neutrons, with each neutron being accompanied by an electron-type neutrino. The neutrinos emerge directly from the core and reach Earth in eight minutes, carrying in their flux and energy distribution a detailed account of the Sun's fusion processes. The neutrino core-shine at Earth is intense (10^{11} neutrinos per square centimetre per second) but difficult to detect, because neutrinos rarely interact with matter as they pass through it. For this reason, the bulk of these neutrinos have evaded direct detection — until now.

All but 1% of solar ^4He synthesis takes place through the *pp* chain, which releases two electron neutrinos and 26.73 million electronvolts of energy. The three cycles of this chain (*pp*I, *pp*II and *pp*III) are each associated with a characteristic neutrino source. All three cycles begin with the fusion of two protons in the solar plasma to form deuterium (^2H), through the '*pp*' and '*pep*' reactions (Fig. 1). The former accounts for 99.76% of deuterium synthesis, and thus determines the rate of solar-energy generation.

In early experiments to detect neutrinos — the historic chlorine experiment⁴, Kamiokande⁵ and GALLEX/SAGE^{6,7} — the number of neutrinos recorded was about one-half to one-third of that predicted by theory, indicating that a basic flaw existed in our understanding of either the Sun or the physics of neutrinos. This solar-neutrino problem motivated the building of a new generation of massive detectors, namely, the Sudbury Neutrino Observatory⁸, Super-Kamiokande⁹ and Borexino (Fig. 2). The Sudbury Neutrino Observatory and Super-Kamiokande, which detected the high-energy neutrinos produced in the β -decay of boron-8 (^8B) in the *pp*III cycle (Fig. 1), traced the problem to new particle physics: neutrinos have a mass and can change 'flavour' during their transit from the solar core to Earth. Two-thirds of solar ^8B electron neutrinos oscillate into other types (muon and tau neutrinos) before reaching Earth.

The ^8B neutrinos comprise about 0.01% of the total flux of neutrinos coming from the Sun. The Borexino detector was designed to detect

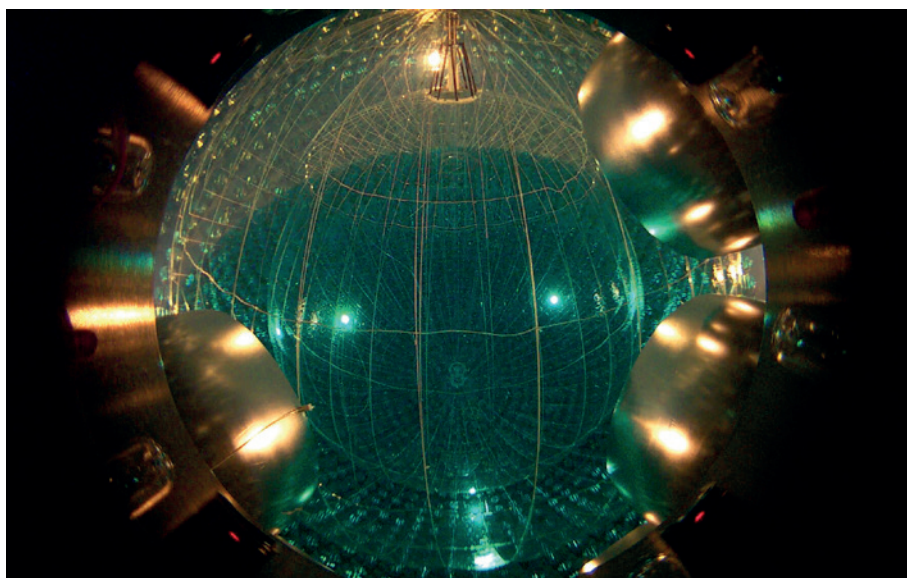


Figure 2 | The Borexino detector. A view inside the Borexino detector showing its nested nylon vessels, as the inner volume is being filled with 278 tonnes of liquid scintillator. Solar neutrinos are detected through their scattering off electrons in the scintillator: the scattered electrons emit scintillation light that is recorded in phototubes. The photo was taken from a camera mounted on the 14-metre-diameter stainless-steel sphere that surrounds the detector. The camera looks out between several of the cones (in the foreground, and thus seemingly enlarged) that help to direct the light into the phototubes.

the remaining, lower-energy solar neutrinos through their scattering off electrons in liquid scintillator: the recoiling electrons emit light that is recorded in the detector. Measurements of the low-energy neutrinos produced in the *pep* reaction and the *pp*II cycle have been announced previously^{10,11}. Now, after an extended detector-purification campaign, the Borexino Collaboration has measured the *pp* neutrinos — the lowest-energy neutrino branch, accounting for 90% of the total flux¹².

This result provides an important test of how matter affects neutrino oscillations — oscillations occurring within the Sun differ from those occurring in a vacuum. Such 'matter effects' can be exploited to determine whether a given neutrino is heavier or lighter than another neutrino. Previous experiments^{8,9} fixed the relative masses of two of the three neutrinos in this way. The theory used predicts¹² that oscillation probabilities will be lower for *pp* neutrinos than for higher-energy ^8B neutrinos: a larger fraction of the *pp* neutrinos will arrive at Earth as electron neutrinos. The fraction found by Borexino, 0.64 ± 0.12 , is nearly twice that found for ^8B neutrinos. This verification of theory is important because future planned 'long-baseline' neutrino-beam experiments will exploit matter effects to determine the ordering among all three neutrino masses.

The Borexino results also provide new tests of the Sun. Most solar-neutrino analyses assume that the total flux of neutrinos is consistent with the solar luminosity. But the connection between neutrino emission, which measures the rate of energy generation in the solar core today, and luminosity is valid only in a steady-state Sun: because photons take

about 100,000 years to diffuse out of the core, this connection would not hold if the temperature of the solar core varies on times less than 100,000 years. Tests of this connection thus constrain solar variability and well as certain new-physics phenomena, such as solar emission of undetected 'sterile' neutrinos. Because the solar luminosity has been measured to a precision of 0.01%, checks on this relationship are limited by neutrino-flux uncertainties. The Borexino Collaboration notes that the 10% uncertainty of its *pp* neutrino-flux determination could be reduced to 1% in an improved experiment.

Now that all four principal neutrino sources from the *pp* chain have been directly measured, one task remains: about 1% of solar ^4He synthesis takes place through the CN cycle. Neutrinos produced through this channel have not yet been detected. Because the CN cycle is catalysed by reactions on C and N, its rate is proportional to the solar core's metallicity (the fraction of elements other than H or ^4He). Consequently, the core metallicity can be deduced from the CN neutrino flux. A measurement of these neutrinos could directly confirm the solar-abundance problem¹²: helioseismic data suggest that the solar interior is metal-rich relative to the solar surface, contradicting the standard-solar-model assumption of a homogeneous primordial Sun, formed from gas that had mixed convectively.

Measuring CN neutrinos is a final challenge for Borexino, requiring further progress in background suppression. One of the candidate explanations¹² for the solar-abundance problem is the effect of planetary formation on the early Sun: this process swept a great deal of metal from gas that, if then deposited on

the Sun's surface, could have diluted the star's outer convective zone. The demonstration of a connection between a host star's metallicity and the presence of its planets would have significant implications for exoplanet searches. ■

Wick Haxton is in the Physics Department, University of California, Berkeley, and the

Nuclear Science Division, Lawrence Berkeley National Laboratory, California 94720, USA. e-mail: haxton@berkeley.edu

1. Borexino Collaboration. *Nature* **512**, 383–386 (2014).
2. Bahcall, J. N. *J. R. Astron. Soc. Can.* **94**, 219–227 (2000).
3. Bethe, H. A. *Phys. Rev.* **55**, 436–456 (1939).
4. Davis, R. Jr, Harmer, D. S. & Hoffman, K. C. *Phys. Rev. Lett.* **20**, 1205–1209 (1968).
5. Fukuda, Y. *et al.* *Phys. Rev. Lett.* **77**, 1683–1686 (1996).

6. GALLEX Collaboration *et al.* *Phys. Lett. B* **447**, 127–133 (1999).
7. Abdurashitov, J. N. *et al.* *Phys. Rev. C* **80**, 015807 (2009).
8. Ahmad, Q. R. *et al.* *Phys. Rev. Lett.* **89**, 011302 (2002).
9. Abe, K. *et al.* *Phys. Rev. D* **83**, 052010 (2011).
10. Bellini, G. *et al.* *Phys. Rev. Lett.* **107**, 141302 (2011).
11. Bellini, G. *et al.* *Phys. Rev. Lett.* **108**, 051302 (2012).
12. Haxton, W. C., Robertson, R. G. H. & Serenelli, A. M. *Annu. Rev. Astron. Astrophys.* **51**, 21–61 (2013).

APPLIED PHYSICS

Hybrid sensors ring the changes

An improved design for a class of magnetometer greatly increases the sensitivity of these devices — and might be the vanguard of a new generation of hybrid sensors that combine different types of signal to increase sensitivity.

JÖRG WRACHTRUP & AMIT FINKLER

When seventeenth-century sailors undertook their dangerous journeys around the Cape of Good Hope or to the Spice Islands, they had an invaluable navigational tool on board: the compass. Only with the aid of this, the most precise measurement device of those times, were they able to accomplish their daring feats. Since then, precision measurements of magnetic fields have been key drivers of basic science and of a whole wealth of technologies with applications ranging from navigation to the medical sciences. Writing in *Advanced Materials*, Forstner *et al.*¹ report a vastly improved design for a certain class of magnetometer — a hybrid sensor that measures magnetic fields using an optical signal.

Various technologies have been developed to measure magnetic fields. The record holders for sensitivity are devices known as atom-vapour magnetometers². Others can measure the dipolar magnetic interaction between two electrons³, and still others, most notably diamond spin sensors, have a compact design that might allow magnetometers to be made as small as one nanometre across. The central element of Forstner and colleagues' sensor is a 'whispering gallery mode' micro-resonator: a doughnut-shaped device that traps light waves and is known for its strong optical confinement and ultra-high quality factor (a measure of the energy lost from the device; high quality factors correspond to low rates of energy loss). The frequency at which light resonates in the doughnut depends on the doughnut's geometry.

To sense magnetic fields, Forstner *et al.* combined the resonator with a magnetostrictive material. This material expands in a magnetic field and thereby exerts a force on the cavity system, changing the micro-resonator's

shape and thus its resonance frequencies. So, by measuring the change in resonance frequencies, the strength of the magnetic field can be determined.

The authors placed the magnetostrictive material in the middle of the doughnut, where it does not interfere with optical measurements and is most effective in changing the shape of the resonator (Fig. 1). Because of the high quality factor of the device, tiny distortions of its shape — and therefore tiny changes of magnetic field — can be detected. In a proof-of-principle experiment, Forstner and co-workers demonstrated that their device has a magnetic-field sensitivity of 100 picotesla (1 picotesla is 10^{-12} tesla) for measurements of 1 second duration. Other kinds of magnetometer, such as atom-vapour types, are sensitive to much lower fields^{4,5} (less than 1 femtotesla; 1 femtotesla is 10^{-15} tesla), but Forstner and co-workers' approach is three orders of magnitude more

sensitive than previous hybrid optomechanical sensors⁶, mostly because of the integration of the magnetostrictive material into the micro-resonator. Moreover, the researchers extended the frequency range of magnetic fields that can be sensitively detected by the sensor by using the nonlinear properties of the magnetostrictive material.

Compared with other, more-sensitive sensor types, the main advantage of Forstner and colleagues' device is its small size, which is limited to the size of the doughnut (about a few tens of micrometres). Furthermore, the device works at room temperature and tolerates background magnetic fields. This opens up various potential applications, such as navigation without the need for a global positioning system, and the detection of weak biomagnetic fields from neurons.

Forstner and co-workers' achievements follow a recent trend in sensor design: the use of optomechanical sensors, in which optical and mechanical signals are combined to enhance sensitivity. By incorporating advances in nanophotonics and the fabrication of nanoscale structures, optomechanical sensors have been made that perform better than their electromechanical counterparts, mostly because the noise in optical readouts is much lower than in electronic ones. Excellent examples include state-of-the-art force sensors⁷ and accelerometers⁸, and a detector of microwave photons⁹.

Hybrid designs based on principles other than optomechanics might also be adopted to

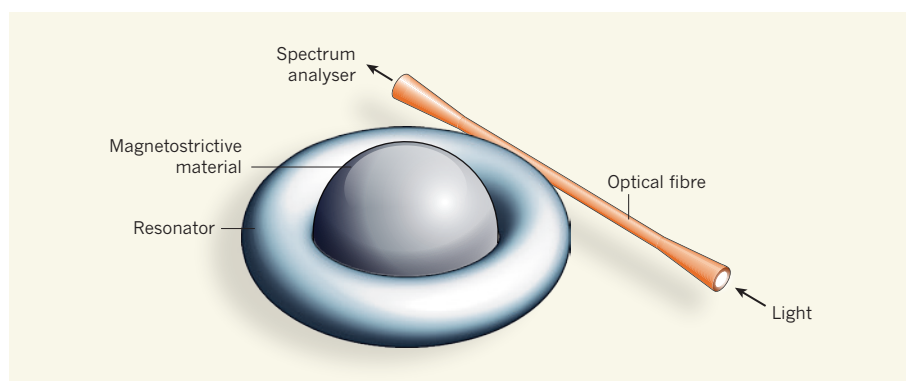


Figure 1 | An optomechanical magnetometer. Forstner *et al.*¹ have constructed a magnetic-field sensor in which a doughnut-shaped 'whispering gallery mode' resonator device surrounds a piece of magnetostrictive material. Light is guided through a tapered optical fibre into the sensor, where it is trapped by the resonator. In the presence of a magnetic field, the magnetostrictive material expands, distorting the resonator and changing the frequency at which light resonates in the device. The change of resonance frequency is measured by a spectrum analyser and used to determine the strength of the magnetic field.

improve the sensitivity of sensors. Consider diamond spin sensors. Rather than using optical resonances, diamond spin sensing relies on the precision measurement of transitions between resonance states that are associated with a quantum property of electrons: spin. Such sensors can be responsive to various quantities, including magnetic¹⁰ and electric fields¹¹, temperature¹², pressure¹³ and strain¹⁴. But the sensitivity of these devices — particularly for temperature and strain — is limited by the nature of the ground-state spin wave function of the electrons used in the sensing mechanism¹³.

It has been proposed¹⁵ that this problem could be overcome by constructing a hybrid sensor using the same magnetostrictive material as that used by Forstner and colleagues. However, the material would function in exactly the opposite way to that reported by these authors: a strain (or an electric field) would generate a magnetic field, which is sensitively detected by the electron spins, which in turn are read out optically. The resulting diamond hybrid sensor is predicted to be about 1,000 times more sensitive for pressure, force or electric field than diamond spins alone, and would retain excellent spatial resolution.

Hybrid sensors thus seem to be an upcoming theme in sensor technology. An important further step will be to use advances in quantum technology to achieve the limits of accuracy.

The resulting quantum hybrid sensors could potentially revolutionize sensor technology in various disciplines, enabling unprecedented opportunities in technology and basic science. ■

Jörg Wrachtrup and Amit Finkler are at the Third Institute of Physics, University of Stuttgart, Stuttgart 70569, Germany.
e-mail: wrachtrup@physik.uni-stuttgart.de

1. Forstner, S. *et al.* *Adv. Mater.* <http://dx.doi.org/10.1002/adma.201401144> (2014).
2. Budker, D. & Romalis, M. *Nature Phys.* **3**, 227–234 (2007).
3. Kotler, S., Akerman, N., Navon, N., Glickman, Y. & Ozeri, R. *Nature* **510**, 376–380 (2014).
4. Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. *Nature* **422**, 596–599 (2003).
5. Lee, S.-K., Sauer, K. L., Seltzer, S. J., Alem, O. & Romalis, V. *Appl. Phys. Lett.* **89**, 214106 (2006).
6. Forstner, S. *et al.* *Phys. Rev. Lett.* **108**, 120801 (2012).
7. Gavartin, E., Verlot, P. & Kippenberg, T. J. *Nature Nanotechnol.* **7**, 509–514 (2012).
8. Krause, A. G., Winger, M., Blasius, T. D., Lin, Q. & Painter, O. *Nature Photon.* **6**, 768–772 (2012).
9. Bagci, T. *et al.* *Nature* **507**, 81–85 (2014).
10. Taylor, J. M. *et al.* *Nature Phys.* **4**, 810–816 (2008).
11. Dolde, F. *et al.* *Nature Phys.* **7**, 459–463 (2011).
12. Neumann, P. *et al.* *Nano Lett.* **13**, 2738–2742 (2013).
13. Doherty, M. W. *et al.* *Phys. Rev. Lett.* **112**, 047601 (2014).
14. Ouartchaiyapong, P., Lee, K. W., Myers, B. A. & Bleszynski Jayich, A. C. *Nature Commun.* **5**, 4429 (2014).
15. Cai, J., Jelezko, F. & Plenio, M. B. *Nature Commun.* **5**, 4065 (2014).

BEHAVIOURAL ECOLOGY

Love thy neighbour

A theoretical model suggests that the cause of female-driven extra-pair mating lies in the spreading of male interests among neighbouring families, creating powerful incentives for male cooperation and concomitant benefits for females.

BEN C. SHELDON & MARC MANGEL

Birds have long served as paragons of contented coupledom, and with apparent justification, given that more than 90% of bird species breed in socially monogamous pairs, with the male and female seemingly cooperating to rear offspring. But appearances can be deceptive. In fact, in the majority of bird species tested — data are available for more than 200 species — a substantial proportion (sometimes exceeding 50%) of offspring are sired by a male other than that providing parental care. A huge, but rather fruitless¹, effort to explain the evolution of female extra-pair mating behaviour in birds has focused on potential genetic benefits to females. Now, writing in *PLoS ONE*, Eliassen and Jørgensen² offer an alternative explanation: that extra-pair mating by females creates an incentive for males to cooperate with their neighbours, generating a

form of collective good that may be a powerful evolutionary force.

The development in the 1980s of genetic markers to assign parentage revolutionized our understanding of mating patterns in many organisms, but nowhere was the effect greater than in birds³. Before this, most birds seemed rather uninteresting from the perspective of sexual selection, because it was expected that social monogamy, reinforced by the need for rapidly growing offspring to receive extensive care from both parents, would limit the extent to which sexual selection could influence fitness (in evolutionary terms, the likelihood of survival and reproductive success). The realization that extra-pair paternity could, in principle, greatly increase the fitness variation among males relative to females, hence creating the potential for strong sexual selection, led to a vigorous interest in testing the causes and consequences of extra-pair copulation



50 Years Ago

‘Obituary: Prof. James Franck’ — I remember his famous lecture ... which gave strong support to Bohr’s new theory of the atom. Einstein said to me: “It’s so lovely, it makes you cry!” ... In 1920 Franck was called to a chair in Göttingen ... many honours ... have come his way, among them the Nobel Prize in Physics in 1926 jointly with Hertz ... Göttingen saw Franck’s happiest and most fruitful period ... Those happy days ended when Hitler came to power. When the racial laws were announced, Franck published a courageous open letter in which he resigned his chair ... He was the most lovable of men because he loved people; kindness shone from his eyes. There must be many beside myself who now feel they have lost one of their best friends.

Lise Meitner

From *Nature* 29 August 1964

100 Years Ago

The Trail of the Sandhill Stag by E. T. Seton; *Wild Game in Zambesia* by R. C. F. Maugham — The first of the books before us aims rather at being a work of art than of natural history ... Never since the days of melancholy Jacques was such an outpour of sentiment upon a stricken deer. To those who like this mood the book may be recommended, for it is curiously wrought and daintily embellished ... The second book on the list strikes quite another strain ... But the list of stores required by two persons for a trip of two months is startling. It almost shakes one’s confidence in the author to learn that he cannot go into camp with a friend for two months without a dozen tins each of lobster and salmon, two dozen tins of sausages, and three dozen tins of fruit in syrup.

From *Nature* 27 August 1914

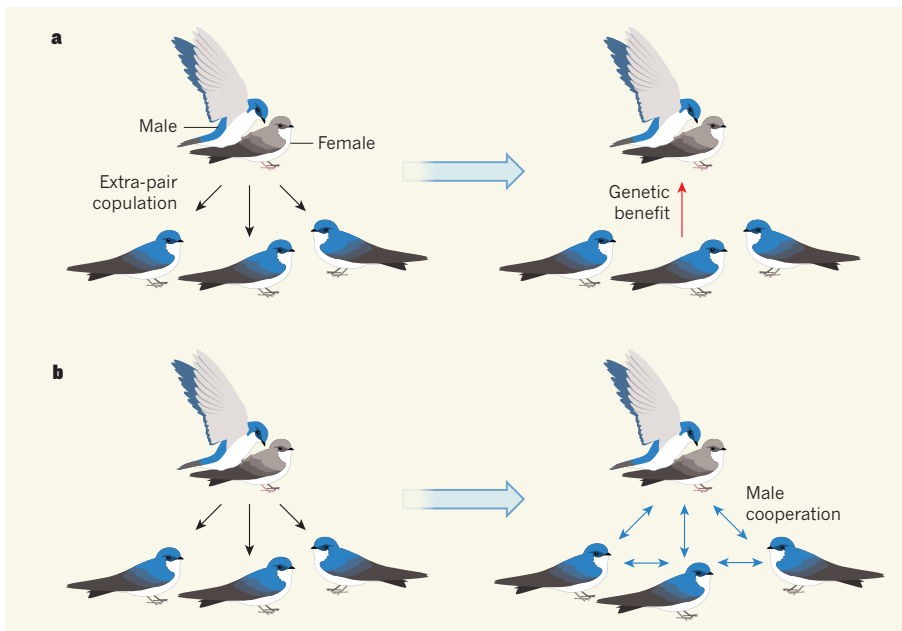


Figure 1 | Potential benefits of extra-pair mating. Genetic analysis of parentage in socially monogamous birds has revealed many cases in which female-driven extra-pair mating leads to males caring for offspring that are not their own. **a**, Previous studies to explain such behaviour focused on the idea that females gain genetic benefits from this behaviour. But the genetic benefit is small (or absent), because it is gained only from extra-pair copulations that result in fertilization. **b**, A new model by Eliassen and Jørgensen² argues that the key force in the evolution of extra-pair mating in birds lies in the way that the matings distribute potential paternity among neighbours and therefore create incentives for cooperation among neighbouring males.

(EPC) behaviour and resulting extra-pair fertilizations.

EPC behaviour in birds is frequently initiated by females, and the only obvious benefit associated with extra-pair fertilizations seemed to be genetic — the provision of genes from the extra-pair partners to offspring. So this seemed an ideal system for testing the ‘good gene’ benefits of sexual selection (Fig. 1a). Over time, however, a large body of empirical work, exemplified by recent comprehensive studies on American song sparrows (*Melospiza melodia*)⁴, led to the conclusion that the genetic benefits of EPCs are at most small. This sparked the suggestion¹ that there has been too much emphasis on adaptive explanations for the evolution of female EPC behaviour, and that non-adaptive explanations, such as inter-sexual genetic constraints (female behaviour resulting from selection in males), should be more rigorously explored. The hypotheses arising from such alternative explanations are difficult to test empirically, although one study⁵ of captive zebra finches was supportive of genetic constraints.

Now, however, Eliassen and Jørgensen use evolutionary modelling to suggest that EPC behaviour in females is indeed adaptive, but that researchers interested in explaining it have simply been looking in the wrong place. The authors’ model (Fig. 1b) argues that female EPC behaviour creates an incentive for males to cooperate with neighbours in a wide range of behaviours that add together to a common good. The incentive occurs because, rather than all of a male’s expected paternity being

limited to a single nest, female EPC spreads the likelihood of paternity among neighbouring nests. Given this uncertain redistribution of a male’s reproductive interests, Eliassen and Jørgensen predict an associated shift by males to invest more in behaviours that benefit neighbouring nests as well.

What form might such behaviours take? In the context of socially monogamous birds, collective behaviour to discourage predators or infanticide by males represents one potentially important mechanism. Equally, this logic might operate through a reduction in territorial or aggressive interactions between neighbours, or by resource sharing during offspring rearing, such as through a relaxation in territorial boundaries. A strength of the new model is that the focus on what type of benefit females gain from extra-pair mating changes from genetic to ecological. Predation and food limitation are commonly identified as contributing to variation in fitness in wild populations, and the effects can be large.

One of the puzzles in understanding EPC behaviour in socially monogamous species is that it occurs in systems that seem likely to be destabilized by such behaviour. Indeed, there is mounting comparative evidence, from social insects⁶ as well as birds⁷, that cooperation is unlikely to evolve within extended families unless there is also genetic monogamy. The perspective offered by Eliassen and Jørgensen is that a different form of cooperation — between unrelated individuals — may actually be promoted by promiscuity. However,

there will inevitably be a trade-off between the group-defence benefits accorded by redistribution of paternity and the expected reduction in a male’s care for offspring in his ‘own’ nest⁸ as his relatedness to these offspring falls; the resolution of this trade-off will probably depend on the relative costs and benefits of the two processes. Thus, their model is as much about the evolution of cooperation as it is about EPC behaviour. Furthermore, because the authors used an adaptive-dynamics method⁹ to build their model, they have laid the groundwork for investigating the broader goal of understanding the emergence of societies as complex adaptive systems¹⁰.

Eliassen and Jørgensen show that many of their model’s predictions, when combined with life-history differences between species, accord well with empirical data from field studies. For instance, the model predicts that extra-pair sires should usually be immediate neighbours, and that exchange of paternity between males will be commonplace — both patterns are commonly seen in the field. Across species, the model explains why extra-pair paternity will be less frequent as longevity increases, a non-intuitive result that accords well with the evidence, but is also predicted by other models. At present, no data exist that could test this model to the exclusion of others, although the model may, as good theory should, challenge empiricists to design more discriminatory tests. Nevertheless, Eliassen and Jørgensen’s work represents a radical shift in focus of tests of the costs and benefits of extra-pair mating in birds, and has broad consequences for our understanding of the evolution of cooperation among interacting, but non-related, individuals. ■

Ben C. Sheldon is at the Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. **Marc Mangel** is at the Center for Stock Assessment Research, University of California, Santa Cruz, Santa Cruz, California 95064, USA. e-mails: ben.sheldon@zoo.ox.ac.uk; msmangel@ucsc.edu

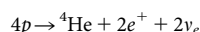
- Forstmeier, W., Nakagawa, S., Griffith, S. C. & Kempenaers, B. *Trends Ecol. Evol.* **29**, 456–464 (2014).
- Eliassen, S. & Jørgensen, C. *PLoS ONE* **9**, e99878 (2014).
- Griffith, S. C., Owens, I. P. F. & Thuman, K. A. *Mol. Ecol.* **11**, 2195–2212 (2002).
- Sardell, R. J., Arcese, P., Keller, L. F. & Reid, J. M. *Am. Nat.* **179**, 779–793 (2012).
- Forstmeier, W., Marin, K., Bolund, E., Schielzeth, H. & Kempenaers, B. *Proc. Natl Acad. Sci. USA* **108**, 10608–10613 (2011).
- Hughes, W. O. H., Oldroyd, B. P., Beekman, M. & Ratnieks, F. L. W. *Science* **320**, 1213–1216 (2008).
- Cornwallis, C. K., West, S. A., Davis, K. E. & Griffin, A. S. *Nature* **466**, 969–972 (2010).
- Sheldon, B. C. *Phil. Trans. R. Soc. Lond. B* **357**, 341–350 (2002).
- Dercole, F. & Rinaldi, S. *Analysis of Evolutionary Processes* (Princeton Univ. Press, 2008).
- Levin, S. *Proc. Natl Acad. Sci. USA* **111** (Suppl. 3), 10781–10784 (2014).

Neutrinos from the primary proton–proton fusion process in the Sun

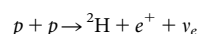
Borexino Collaboration*

In the core of the Sun, energy is released through sequences of nuclear reactions that convert hydrogen into helium. The primary reaction is thought to be the fusion of two protons with the emission of a low-energy neutrino. These so-called *pp* neutrinos constitute nearly the entirety of the solar neutrino flux, vastly outnumbering those emitted in the reactions that follow. Although solar neutrinos from secondary processes have been observed, proving the nuclear origin of the Sun's energy and contributing to the discovery of neutrino oscillations, those from proton–proton fusion have hitherto eluded direct detection. Here we report spectral observations of *pp* neutrinos, demonstrating that about 99 per cent of the power of the Sun, 3.84×10^{33} ergs per second, is generated by the proton–proton fusion process.

We have known for 75 years that the energy generated by stars comes from the fusion of light nuclei into heavier ones^{1–3}. In the Sun, hydrogen is transformed into helium predominantly via the *pp* cycle^{4,5}, a chain of reactions releasing 26.73 MeV and electron neutrinos ν_e , and summarized as



The cycle begins with the fusion of two protons into a deuteron, which occurs 99.76% of the time⁶ by means of the primary reaction



Neutrinos produced in this step are referred to as *pp* neutrinos. Some of the nuclear reactions that follow also produce neutrinos of various energies. ${}^4\text{He}$ may also be formed through the CNO (carbon–nitrogen–oxygen) cycle², which is thought to be predominant in heavy stars, but to produce at most 1% of the Sun's energy^{7,8}. Present models of the Sun⁹ precisely predict the flux and energy distribution of emitted neutrinos (Fig. 1). So far, only the radiochemical gallium experiments (after the first observation by GALLEX^{10,11} and, later, by SAGE¹²) have been sensitive to *pp* solar neutrinos ($0 < E < 420$ keV). However, by measuring only an integrated flux of all solar electron neutrinos above an energy threshold (233 keV), the *pp* neutrino flux could be extracted only indirectly, by combining the GALLEX and SAGE measurements with those of other experiments^{13–17}.

The Borexino experiment came online in 2007 with high sensitivity to all solar neutrino components, particularly those below 2 MeV (Fig. 1). Borexino has made the first measurement of ${}^7\text{Be}$ neutrinos¹⁷ and proton–electron–proton (*pep*) neutrinos¹⁸, measured ${}^8\text{B}$ neutrinos¹⁹ at a lower energy threshold than other experiments, and set the best available limit on the solar CNO neutrino component¹⁸. The detection of *pep* neutrinos itself indirectly indicates the existence of *pp* neutrinos, because the $p + e + p \rightarrow {}^2\text{H} + \nu_e$ reaction is a rare (0.24%; ref. 6) alternative first step of the *pp* cycle. Attempts to measure *pp* neutrinos directly over the past 30 years (see ref. 20 for a recent review) have been hindered by the inability to sufficiently suppress radioactive backgrounds in this low-energy region. The Borexino detector, which is designed to minimize backgrounds from radioactive isotopes both within, and external to, the liquid scintillator target, made it possible to search for the very low-energy *pp*

neutrinos. The measured solar *pp* neutrino flux is $(6.6 \pm 0.7) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$, in good agreement with the prediction of the standard solar model⁹ (SSM) $(5.98 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1})$.

The observation of *pp* neutrinos provides us with a direct glimpse at the keystone fusion process that keeps the Sun shining and strongly reinforces our theories on the origin of almost the entirety of the Sun's energy. Their measured flux can also be used to infer the total energy radiated by the Sun, $3.84 \times 10^{33} \text{ erg s}^{-1}$. However, because photons produced in the Sun's core take a very long time (at least a hundred thousand years; ref. 21) to reach the surface, neutrino and optical observations in combination provide experimental confirmation that the Sun has been in thermodynamic equilibrium over such a timescale.

Searching for *pp* neutrinos with Borexino

The Borexino experiment (Methods) detects solar neutrinos by measuring the energy deposited in the liquid scintillator target by recoiling electrons undergoing neutrino–electron elastic scattering:

$$\nu_x + e \rightarrow \nu_x + e \quad (1)$$

where x denotes one of the three neutrino flavours (e , μ , τ). The detector is fully described in ref. 22.

The solar neutrino flux reaching the Earth is composed not only of electron neutrinos produced in the nuclear reactions in the Sun, but, owing to the process of flavour oscillations (Methods), also of muon and tau neutrinos. The *pp* neutrino energy spectrum extends up to 420 keV, yielding a maximum electron recoil energy of $E_{\text{max}} = 264$ keV (ref. 23). The expected flux of *pp* neutrinos is calculated in the framework of the SSM (Methods). The most recent calculations are those of ref. 9 (other models, such as the one described in ref. 24, give similar results). The predictions for the total flux of *pp* neutrinos at Earth⁹ range between $5.98 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$, for the high-metallicity model, and $6.03 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$, for the low-metallicity model. The latest values of the neutrino oscillation parameters²⁵ are needed to calculate the relative proportions of the three flavours within the solar neutrino flux at Earth. For reference, when combining them with the high-metallicity SSM prediction (assumed throughout this paper unless otherwise specified) and using neutrino–electron scattering cross-sections derived from refs 25, 26 (also P. Langacker and J. Erler, personal communication), we expect the *pp* neutrino interaction

*Lists of participants and their affiliations appear at the end of the paper.

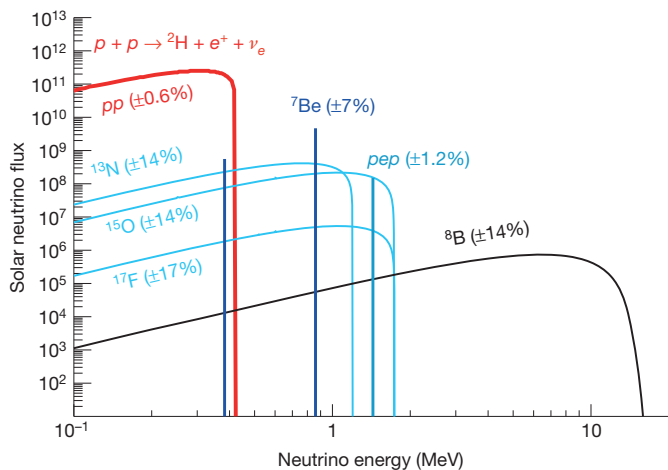


Figure 1 | Solar neutrino energy spectrum. The flux (vertical scale) is given in $\text{cm}^{-2} \text{s}^{-1} \text{MeV}^{-1}$ for continuum sources and in $\text{cm}^{-2} \text{s}^{-1}$ for mono-energetic ones. The quoted uncertainties are from the SSM⁹.

(equation (1)) rate to be 131 ± 2 counts per day (c.p.d.) per 100 t of target scintillator.

The scintillation light generated by a 100 keV event typically induces signals in ~ 50 photomultiplier tubes (PMTs). This allows for a low detection threshold (~ 50 keV), much less than the maximum electron recoil energy of pp neutrinos ($E_{\text{max}} = 264$ keV).

The pp neutrino analysis is performed through a fit of the energy distribution of events selected to maximize the signal-to-background ratio. The selection criteria (Methods) remove residual cosmic muons, decays of muon-produced isotopes, and electronic noise events. Furthermore, to suppress background radiation from external detector components, only events whose position is reconstructed inside the central detector volume (the ‘fiducial volume’: 86 m^3 , 75.5 t) are used in the analysis. The fit is done within a chosen energy interval and includes all relevant solar neutrino components and those from various backgrounds, mostly from residual radioactivity traces dissolved in the scintillator.

Figure 2 shows a calculation of the spectral shape of the pp neutrino signal (thick red line), as well as of the other solar neutrino components (^7Be , pep and CNO), and of the relevant backgrounds (^{14}C , intrinsic to the organic liquid scintillator; its ‘pile-up’ (see definition below); ^{210}Bi ; ^{210}Po ; ^{85}Kr ; and ^{214}Pb), all approximately at the observed rates in the data. The pp neutrino spectral component is clearly distinguished from those of ^{85}Kr , ^{210}Bi , CNO and ^7Be , all of which have flat spectral shapes in the energy region of the fit. Most of the pp neutrino events are buried

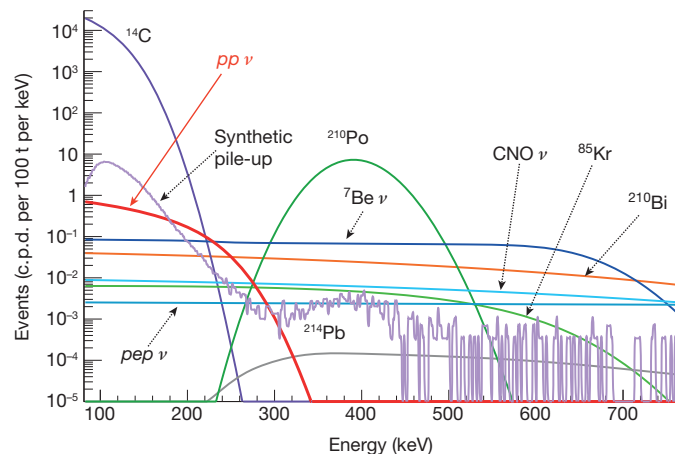


Figure 2 | Energy spectra for all the solar neutrino and radioactive background components. All components are obtained from analytical expressions, validated by Monte Carlo simulations, with the exception of the synthetic pile-up, which is constructed from data (see text for details).

under the vastly more abundant ^{14}C , which is a β -emitter with a Q value of 156 keV. In spite of its tiny isotopic fraction in the Borexino scintillator ($^{14}\text{C}/^{12}\text{C} \approx 2.7 \times 10^{-18}$), ^{14}C β -decay is responsible for most of the detector triggering rate (~ 30 counts s^{-1} at our chosen trigger threshold). The ^{14}C and pp neutrino energy spectra are, however, distinguishable in the energy interval of interest.

The ^{14}C rate was determined independently from the main analysis, by looking at a sample of data in which the event causing the trigger is followed by a second event within the acquisition time window of $16 \mu\text{s}$. This second event, which is predominantly due to ^{14}C , does not suffer from hardware trigger-threshold effects and can thus be used to study the rate and the spectral shape of this contaminant. We measure a ^{14}C rate of 40 ± 1 Bq per 100 t. The error accounts for systematic effects due to detector response stability in time, uncertainty in the ^{14}C spectral shape²⁷, and fit conditions (Methods).

An important consideration in this analysis were the pile-up events: occurrences of two uncorrelated events so closely in time that they cannot be separated and are measured as a single event. Figure 2 shows the expected pile-up spectral shape, which is similar to that of the pp neutrinos. Fortunately, the pile-up component can be determined independently, using a data-driven method, which we call ‘synthetic pile-up’ (Methods). This method provides the spectral shape and the rate of the pile-up component, and is constructed as follows. Real triggered events without any selection cuts are artificially overlapped with random data samples. The combined synthetic events are selected and reconstructed using the same procedure applied to the regular data. Thus, some systematic effects, such as the position reconstruction of pile-up events, are automatically taken into account. The synthetic pile-up is mainly due to the overlap of two ^{14}C events, but includes all possible event combinations, for example ^{14}C with the external background, PMT dark noise or ^{210}Po . ^{14}C – ^{14}C events dominate the synthetic pile-up spectrum between approximately 160 and 265 keV. The fit to the ^{14}C – ^{14}C pile-up analytical shape in this energy region gives a total rate for ^{14}C – ^{14}C pile-up events of 154 ± 10 c.p.d. per 100 t in the whole spectrum, without threshold.

Measurement of the pp neutrino flux

The data used for this analysis were acquired from January 2012 to May 2013 (408 days of data; Borexino Phase 2). This is the purest data set available, and was obtained after an extensive purification campaign that was performed in 2010 and 2011²⁸ and reduced, in particular, the content of ^{85}Kr and ^{210}Bi isotopes, which are important backgrounds in the low-energy region.

The pp neutrino rate has been extracted by fitting the measured energy spectrum of the selected events in the 165–590 keV energy window with the expected spectra of the signal and background components. The energy scale in units of kiloelectronvolts is determined from the number of struck PMTs, using a combination of calibration data collected with radioactive sources deployed inside the scintillator²⁹ and a detailed Monte Carlo model²⁸.

The fit is done with a software tool developed for previous Borexino measurements²⁸ and improved for this analysis to include the description of the response of the scintillator to mono-energetic electrons, to give high statistics; a modified description of the scintillation line-width at low energy, providing the appropriate response functions widths for α - and β -particles (mainly from the ^{210}Po and ^{14}C backgrounds); and the introduction of the synthetic pile-up.

The main components of the fit are the solar neutrino signal (the dominant pp component and the low-energy parts of the ^7Be , pep and CNO components); the dominant ^{14}C background and the associated pile-up; and other identified radioactive backgrounds (^{85}Kr , ^{210}Bi , ^{210}Po and ^{214}Pb). The free fit parameters are the rates of the pp solar neutrinos and of the ^{85}Kr , ^{210}Bi and ^{210}Po backgrounds. The ^7Be neutrino rate is constrained at the measured value¹⁷ within the error, and pep and CNO neutrino contributions are fixed at the levels of the SSM⁹, taking into account the values of the neutrino oscillation parameters²⁵. The ^{14}C and the synthetic pile-up rates are determined from the data

independently and fixed in the fit, allowing for a variation consistent with their measured uncertainty. The ^{214}Pb rate is fixed by the measured rate of fast, time-correlated $^{214}\text{Bi}(\beta)$ – $^{214}\text{Po}(\alpha)$ coincidences. The scintillator light yield and two energy resolution parameters are left free in the fit.

The energy spectrum with the best-fit components is shown in Fig. 3. The corresponding values of the fitted parameters are given in Table 1.

Many fits have been performed with slightly different conditions to estimate the robustness of the analysis procedure. In particular, we varied the energy estimator, the fit energy range, the data selection criteria and the pile-up evaluation method (Methods). The root mean square of the distribution of all the fits is our best estimate of the systematic error (7%). In addition, a systematic uncertainty (2%) due to the nominal fiducial mass determination is added in quadrature; this was obtained from calibration data by comparing the reconstructed and nominal positions of a (^{222}Rn – ^{14}C) radioactive source located near the border of the fiducial volume²⁹. Other possible sources of systematic errors, like the dependence of the result on the details of the energy scale definition and on the uncertainties in the ^{14}C and ^{210}Bi β -decay shape factors, were investigated and found to be negligible (Methods). We also verified that varying the pp and CNO neutrino rates within the measured or theoretical uncertainties changed the pp neutrino rate by less than 1%. We finally confirmed that the fit performed without constraining the ^{14}C rate returns a ^{14}C value consistent with the one previously measured independently (see above) and does not affect the pp neutrino result. The systematic errors are given in Table 1 for all fitted species.

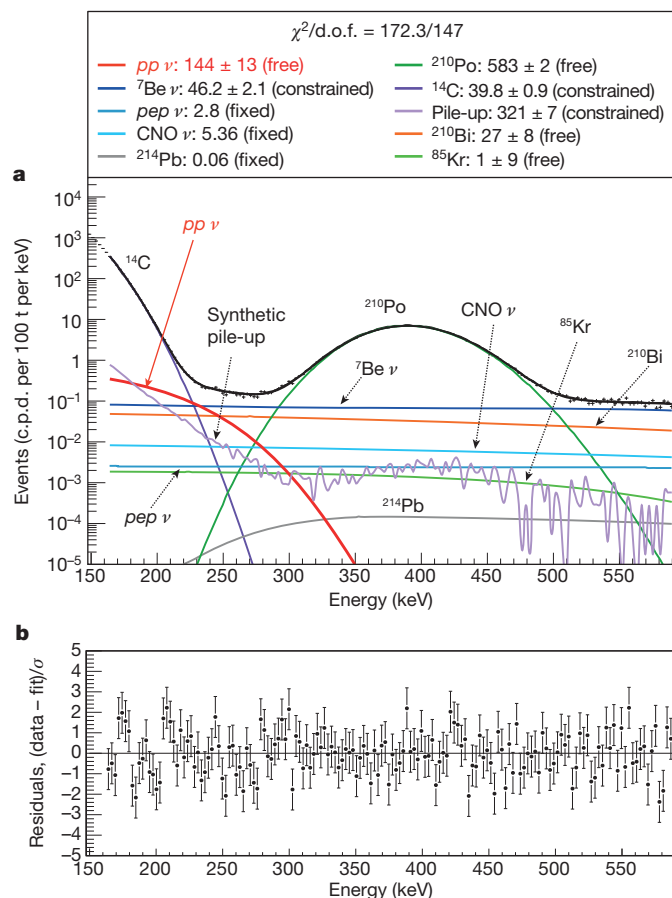


Figure 3 | Fit of the energy spectrum between 165 and 590 keV. **a**, The best-fit pp neutrino component is shown in red, the ^{14}C background in dark purple and the synthetic pile-up in light purple. The large green peak is ^{210}Po α -decays. ^{7}Be (dark blue), pep and CNO (light blue) solar neutrinos, and ^{210}Bi (orange) are almost flat in this energy region. The values of the parameters (in c.p.d. per 100 t) are in the inset above the figure. **b**, Residuals. Error bars, 1σ .

Table 1 | Results from the fit to the energy spectrum

Parameter	Rate \pm statistical error (c.p.d. per 100 t)	Systematic error (c.p.d. per 100 t)
pp neutrino	144 ± 13	± 10
^{85}Kr	1 ± 9	± 3
^{210}Bi	27 ± 8	± 3
^{210}Po	583 ± 2	± 12

The best-fit value and statistical uncertainty for each component are listed together with its systematic error. The χ^2 per degree of freedom of the fit is $\chi^2/\text{d.o.f.} = 172.3/147$.

We note that the very low ^{85}Kr rate (Table 1) is consistent with the independent limit (<7 c.p.d. per 100 t, 95% confidence level) obtained by searching for the β – γ delayed coincidence $^{85}\text{Kr} \rightarrow ^{85\text{m}}\text{Rb} \rightarrow ^{85}\text{Rb}$ (lifetime of the intermediate metastable isotope, $\tau = 1.46 \mu\text{s}$; branching ratio, 0.43%).

We have checked for possible residual backgrounds generated by nuclear spallation processes produced by cosmic ray muons that interact in the detector. We detect these muons with $>99.9\%$ efficiency³⁰. We increased the time window for the muon veto from 300 ms to 5 s and observed no difference in the results. Furthermore, we searched for other possible background due to radioisotopes with sizeable natural abundances and sufficiently long half-lives to survive inside the detector over the timescale of this measurement. These include low-energy α -emitters such as ^{222}Rn and ^{218}Po (both belonging to the radon decay chain), ^{147}Sm and ^{148}Sm , and β -emitters (^7Be), which are all estimated to be negligible and are excluded from the final fit. One β -emitter, ^{87}Rb (half-life, $t_{1/2} = 4.7 \times 10^{10}$ yr; 28% isotopic abundance; $Q = 283.3$ keV), is of particular concern because of the relatively high abundance of Rb in the Earth's crust. Rubidium is an alkali chemically close to potassium but typically 2,000–4,000 times less abundant in the crust. Under these assumptions, and using the measured ^{40}K ($t_{1/2} = 0.125 \times 10^{10}$ yr; 0.0117% isotopic abundance) activity in the fiducial volume, that is, <0.4 c.p.d. per 100 t at the 95% confidence level¹⁸, the ^{87}Rb activity in the Borexino scintillator can be constrained to be much less than 0.1 c.p.d. per 100 t, which is negligible for this analysis. A deviation from the crustal isotopic ratio by a factor of 100 would still keep this background at ~ 1 c.p.d. per 100 t.

The solar pp neutrino interaction rate measured by Borexino is 144 ± 13 (stat.) ± 10 (syst.) c.p.d. per 100 t. The stability and robustness of the measured pp neutrino interaction rate was verified by performing fits with a wide range of different initial conditions. The absence of pp solar neutrinos is excluded with a statistical significance of 10σ (Methods). Once statistical and systematic errors are added in quadrature and the latest values of the neutrino oscillation parameters²⁵ are taken into account, the measured solar pp neutrino flux is $(6.6 \pm 0.7) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$. This value is in good agreement with the SSM prediction⁹ $(5.98 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1})$. It is also consistent with the flux calculated by performing a global analysis of all existing solar neutrino data, including the ^8B , ^7Be and pep fluxes and solar neutrino capture rates^{31,32}. Finally, the probability that pp neutrinos produced in the core of the Sun are not transformed into muon or tau neutrinos by the neutrino oscillation mechanism is found to be $P(\nu_e \rightarrow \nu_e) = 0.64 \pm 0.12$, providing a constraint on the Mikheyev–Smirnov–Wolfenstein large-mixing-angle (MSW-LMA) solution^{25,33,34} in the low-energy vacuum regime (Methods).

Outlook

The proton–proton fusion reaction in the core of the Sun is the keystone process for energy production in the Sun and in Sun-like stars. The observation of the low-energy (0–420 keV) pp neutrinos produced in this reaction was possible because of the unprecedentedly low level of radioactivity reached inside the Borexino detector. The measured value is in very good agreement with the predictions of both the high-metallicity and the low-metallicity SSMs. Although the experimental uncertainty does not yet allow the details of these models to be distinguished, this measurement strongly confirms our understanding of the Sun. Future Borexino-inspired experiments might be able to measure solar pp neutrinos with the level of precision ($\sim 1\%$) needed to cross-compare photon

and neutrino solar luminosities, while providing insight into solar dynamics over 10^5 -yr timescales. At the same time, such a precise measurement of pp neutrinos would yield the ultimate test for the MSW-LMA neutrino oscillation model and allow precision tests for exotic neutrino properties³⁵.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 April; accepted 18 July 2014.

- Bethe, H. A. & Critchfield, C. L. The formation of deuterons by proton combination. *Phys. Rev.* **54**, 248–254 (1938).
- Bethe, H. A. Energy production in stars. *Phys. Rev.* **55**, 434–456 (1939).
- von Weizsäcker, C. F. Über Elementumwandlungen im Innern der Sterne. *Phys. Z.* **38**, 176–191 (1937).
- Fowler, W. A. Completion of the proton-proton reaction chain and the possibility of energetic neutrino emission by hot stars. *Astrophys. J.* **127**, 551–556 (1958).
- Cameron, A. W. G. Nuclear astrophysics. *Annu. Rev. Nucl. Sci.* **8**, 299–326 (1958).
- Adelberger, E. G. *et al.* Solar fusion cross sections II: the pp chain and CNO cycles. *Rev. Mod. Phys.* **83**, 195–245 (2011).
- Bahcall, J. N., Gonzalez-Garcia, M. C., & Peña-Garay, C. Does the Sun shine by pp or CNO fusion reactions? *Phys. Rev. Lett.* **90**, 131301 (2003).
- Haxton, W. & Serenelli, A. CN-cycle solar neutrinos and the Sun's primordial core metallicity. *Astrophys. J.* **687**, 678–691 (2008).
- Serenelli, A. M., Haxton, W. C. & Peña-Garay, C. Solar models with accretion - I. Application to the solar abundance problem. *Astrophys. J.* **743**, 24 (2011).
- Anselmann, P. *et al.* (GALLEX Collaboration). Solar neutrinos observed by GALLEX at Gran Sasso. *Phys. Lett. B* **285**, 376–389 (1992).
- Krauss, L. M. Reopening the solar neutrino question. *Nature* **357**, 437 (1992).
- Abdurashitov, J. N. *et al.* Results from SAGE (the Russian-American solar neutrino gallium experiment). *Phys. Lett. B* **328**, 234–248 (1994).
- Cleveland, B. T. *et al.* Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *Astrophys. J.* **496**, 505–526 (1998).
- Fukuda, Y. *et al.* Solar neutrino data covering solar cycle 22. *Phys. Rev. Lett.* **77**, 1683–1686 (1996).
- Abe, Y. *et al.* (SuperKamioke Collaboration). Solar neutrino results in SuperKamioke-III. *Phys. Rev. D* **83**, 052010 (2011).
- Ahmad, Q. R. *et al.* (SNO Collaboration). Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by 8B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.* **87**, 071301 (2001).
- Bellini, G. *et al.* (Borexino Collaboration). Precision measurement of the 7Be solar neutrino interaction rate in Borexino. *Phys. Rev. Lett.* **107**, 141302 (2011).
- Bellini, G. *et al.* (Borexino Collaboration). First evidence of pep solar neutrinos by direct detection in Borexino. *Phys. Rev. Lett.* **108**, 051302 (2012).
- Bellini, G. *et al.* (Borexino Collaboration). Measurement of the solar 8B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector. *Phys. Rev. D* **82**, 033006 (2010).
- Donald, A. SNO and future solar neutrino experiments. *Nucl. Phys. B Proc. Suppl.* **235–236**, 61–67 (2013).
- Mitalas, R. & Sills, K. R. On the photon diffusion time scale for the Sun. *Astrophys. J.* **401**, 759–760 (1992).
- Alimonti, G. *et al.* The Borexino detector at the Laboratori Nazionali del Gran Sasso. *Nucl. Instrum. Methods A* **600**, 568–593 (2009).
- Bahcall, J. N. Gallium solar neutrino experiments: absorption cross sections, neutrino spectra, and predicted event rates. *Phys. Rev. C* **56**, 3391–3409 (1997).
- Turck-Chièze, S. *et al.* Solar neutrinos, helioseismology and the solar internal dynamics. *Rep. Prog. Phys.* **74**, 086901 (2011).
- Beringer, J. *et al.* (Particle Data Group). Review of particle physics. *Phys. Rev. D* **86**, 010001 (2012); and 2013 partial update for the 2014 edition.
- Bahcall, J. N., Kamionkowski, M. & Sirlin, A. Solar neutrinos: radiative corrections in neutrino-electron scattering experiments. *Phys. Rev. D* **51**, 6146–6158 (1995).
- Kuzminov, V. V. & Osetrova, N. Ja. Precise measurement of ^{14}C beta spectrum by using a wall-less proportional counter. *Phys. At. Nucl.* **63**, 1292–1296 (2000).
- Bellini, G. *et al.* (Borexino Collaboration). Final results of Borexino Phase-I on low-energy solar neutrino spectroscopy. *Phys. Rev. D* **89**, 112007 (2014).
- Back H. *et al.* Borexino calibrations: hardware, methods and results. *J. Instrum.* **7**, P10018 (2012).
- Bellini, G. *et al.* Muon and cosmogenic neutron detection in Borexino. *J. Instrum.* **6**, P05005 (2012).
- Arpesella, C. *et al.* (Borexino Collaboration). Direct measurement of the 7Be solar neutrino flux with 192 days of Borexino data. *Phys. Rev. Lett.* **101**, 091302 (2008).

- Ianni, A. Solar neutrinos and the solar model. *Phys. Dark Univ.* **4**, 44–49 (2014).
- Wolfenstein, L. Neutrino oscillations in matter. *Phys. Rev. D* **17**, 2369–2374 (1978).
- Mikheyev, S. P. & Smirnov, A. Yu. Resonant amplification of neutrino oscillations in matter and spectroscopy of solar neutrinos. *Sov. J. Nucl. Phys.* **42**, 913–917 (1985).
- Haxton, W. C., Robertson, R. G. H. & Serenelli, A. M. Solar neutrinos: status and prospects. *Annu. Rev. Astron. Astrophys.* **51**, 21–61 (2013).

Acknowledgements The Borexino program is made possible by funding from the INFN (Italy); the NSF (USA); the BMBF, DFG and MPG (Germany); the JINR; the RFBR, RSC and NRC Kurchatov Institute (Russia); and the NCN (Poland). We acknowledge the support of the Laboratori Nazionali del Gran Sasso (Italy).

Author Contributions The Borexino detector was designed, constructed and commissioned by the Borexino Collaboration over the span of more than 15 years. The Borexino Collaboration sets the science goals. Scintillator purification and handling, source calibration campaigns, PMT and electronics operations, signal processing and data acquisition, Monte Carlo simulations of the detector, and data analyses were performed by Borexino Collaboration members, who also discussed and approved the scientific results. The manuscript was prepared by a subgroup of authors appointed by the collaboration and subject to an internal collaboration-wide review process. All authors reviewed and approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.S. (borex-spokeperson@lngs.infn.it).

Borexino Collaboration G. Bellini¹, J. Benziger², D. Bick³, G. Bonfini⁴, D. Bravo⁵, B. Caccianiga¹, L. Cadonati⁶, F. Calaprice^{7,8}, A. Caminata⁹, P. Cavalcante⁴, A. Chavarria⁷, A. Chepurinov¹⁰, D. D'Angelo¹, S. Davini¹¹, A. Derbin¹², A. Empl¹¹, A. Etenko^{13,14}, K. Fomenko^{4,15}, D. Franco¹⁶, F. Gabriele⁴, C. Galbiati⁷, S. Gazzana⁴, C. Ghiano⁹, M. Giammarchi¹, M. Göger-Neff¹⁷, A. Goretti⁷, M. Gromov¹⁰, C. Hagner³, E. Hungerford¹, Aldo Ianni⁴, Andrea Ianni⁷, V. Kobaychev¹⁸, D. Korabiev¹⁵, G. Korga¹¹, D. Kryn¹⁶, M. Laubenstein⁴, B. Lehnert¹⁹, T. Lewke¹⁷, E. Litvinovich^{13,14}, F. Lombardi⁴, P. Lombardi¹, L. Ludhova¹, G. Lukyanenko¹³, I. Machulin^{13,14}, S. Manecki⁵, W. Maneschg²⁰, S. Marocchi^{8,9}, Q. Meindl¹⁷, E. Meroni¹, M. Meyer³, L. Miramonti¹, M. Misiaszek²¹, M. Montuschi⁴, P. Mosteiro⁷, V. Muratova¹², L. Oberauer¹⁷, M. Obolensky¹⁶, F. Orlica²², K. Otis⁶, M. Pallavicini⁹, L. Papp^{5,17}, L. Perasso⁹, A. Pocar⁶, G. Ranucci¹, A. Razeto⁴, A. Re¹, A. Romani²², N. Rossi⁴, R. Saldanha⁷, C. Salvo⁹, S. Schönert¹⁷, H. Simgen²⁰, M. Skorokhvatov^{13,14}, O. Smirnov¹⁵, A. Sotnikov¹⁵, S. Sukhotin¹³, Y. Suvorov^{4,13,23}, R. Tartaglia⁴, G. Testera⁹, D. Vignaud¹⁶, R. B. Vogelaar⁵, F. von Feilitzsch¹⁷, H. Wang²³, J. Winter²⁴, M. Wojcik²¹, A. Wright⁷, M. Wurm²⁴, O. Zaimidoroga¹⁵, S. Zavatarelli⁹, K. Zuber¹⁹ & G. Zuzel²¹

Affiliations for participants: ¹Dipartimento di Fisica, Università degli Studi e INFN, 20133 Milano, Italy. ²Chemical Engineering Department, Princeton University, Princeton, New Jersey 08544, USA. ³Institut für Experimentalphysik, Universität Hamburg, 22761 Hamburg, Germany. ⁴INFN Laboratori Nazionali del Gran Sasso, 67100 Assergi, Italy. ⁵Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA. ⁶Physics Department, University of Massachusetts, Amherst, Massachusetts 01003, USA. ⁷Physics Department, Princeton University, Princeton, New Jersey 08544, USA. ⁸Gran Sasso Science Institute (INFN), 67100 L'Aquila, Italy. ⁹Dipartimento di Fisica, Università degli Studi e INFN, 16146 Genova, Italy. ¹⁰Lomonosov Moscow State University Skobeltsyn Institute of Nuclear Physics, 119234 Moscow, Russia. ¹¹Department of Physics, University of Houston, Houston, Texas 77204, USA. ¹²St Petersburg Nuclear Physics Institute, 188350 Gatchina, Russia. ¹³NRC Kurchatov Institute, 123182 Moscow, Russia. ¹⁴National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409 Moscow, Russia. ¹⁵Joint Institute for Nuclear Research, 141980 Dubna, Russia. ¹⁶APC, Université Paris Diderot, CNRS/IN2P3, CEA/IRFU, Observatoire de Paris, Sorbonne Paris Cité, 75205 Paris Cedex 13, France. ¹⁷Physik-Department and Excellence Cluster Universe, Technische Universität München, 85748 Garching, Germany. ¹⁸Kiev Institute for Nuclear Research, 03680 Kiev, Ukraine. ¹⁹Department of Physics, Technische Universität Dresden, 01062 Dresden, Germany. ²⁰Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany. ²¹M. Smoluchowski Institute of Physics, Jagiellonian University, 30059 Krakow, Poland. ²²Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi e INFN, 06123 Perugia, Italy. ²³Physics and Astronomy Department, University of California Los Angeles (UCLA), Los Angeles, California 90095, USA. ²⁴Institut für Physik, Johannes Gutenberg Universität Mainz, 55122 Mainz, Germany.

METHODS

The Borexino detector. The detector (Extended Data Fig. 1) is located deep underground (3,800 m of water equivalent) at the Gran Sasso laboratory, in central Italy²². The active neutrino target, 278 t of ultrapure liquid scintillator (pseudocumene (1,2,4-trimethylbenzene) solvent with 1.5 g l^{-1} 2,5-diphenyloxazole (PPO) wavelength-shifting fluor), is contained inside a thin transparent nylon spherical vessel of 8.5 m diameter. Borexino detects solar neutrinos by measuring the energy deposited by recoiling electrons following neutrino–electron elastic scattering. The scintillator promptly converts the kinetic energy of electrons into photons, detected and converted into electronic signals (photoelectrons) by 2,212 PMTs mounted on a concentric 13.7 m-diameter stainless steel sphere (SSS). The interaction in the scintillator of ionizing particles, such as those resulting from radioactive decays inside the detector (that we call background), can mimic and cover the expected signal. Every effort was made to minimize radioactive contamination of the scintillator and all surrounding detector materials. (1) The scintillator was produced from crude oil with minimal amounts of ^{14}C isotopic contamination (a key feature for this particular analysis). (2) Borexino has developed specific purification techniques^{36,37} for scintillator, water and all other detector components, implemented innovative cleaning procedures and operated a 4 t prototype, the Counting Test Facility³⁸; all this made it possible to reduce the radioactive contamination in ^{238}U and ^{232}Th to less than 10^{-18} g per gram of liquid scintillator, a level never reached before²². The Counting Test Facility also proved that organic liquid scintillator produced from ancient and deep underground oil sources can yield a very low $^{14}\text{C}/^{12}\text{C}$ ratio³⁹. (3) The volume between the nylon vessel and the SSS is filled with 889 t of ultrapure non-scintillating fluid, and acts as radiation shield for radioactivity emitted by the PMTs. A second, larger (11.5 m diameter) nylon sphere prevents radon and other radioactive contaminants from the PMTs and SSS from diffusing into the inner part of the detector. The sphere is immersed in a 2,100 t water Cherenkov detector for residual cosmic muons, which can induce background via spallation processes with the scintillator.

Solar electron neutrino survival probability. After the SNO experiment in 2001¹⁶, the deficit of solar ν_e observed on Earth has been explained by the neutrino oscillation mechanism: these ν_e undergo lepton flavour transformation into ν_μ or ν_τ in a quantum mechanical process requiring neutrinos to have a finite mass difference. The exact phenomenon is somewhat more complex (see ref. 35 for a recent review): beyond the neutrino oscillation mechanism in vacuum itself, there is a specific feature coming from matter effects while the neutrinos travel inside the Sun (the MSW effect^{33,34}). This effect is energy dependent and is fully effective for solar neutrinos above ~ 5 MeV. Below 1 MeV, vacuum oscillations dominate, and a smooth transition region is predicted between ~ 1 and ~ 5 MeV. One of the parameters of oscillation is one of the three mixing angles between the neutrino species, for which all experimental data (from solar neutrino or reactor antineutrino experiments) isolate one particular solution, called the large-mixing-angle (LMA) solution; the model describing the solar neutrino results is now called MSW-LMA. The Borexino data, which cover the full solar neutrino spectrum, make it possible to test the predictions of the MSW-LMA model in all oscillation regimes with the same detector. Extended Data Fig. 2 shows the probability $P(\nu_e \rightarrow \nu_e)$ that ν_e produced in the core of the Sun are not transformed into ν_μ or ν_τ via the neutrino oscillation mechanism. The violet band corresponds to the 1σ prediction of the MSW-LMA solution, using the most recent values for the oscillation parameters²⁵. The survival probability for the pp neutrino data from the present paper is $P(\nu_e \rightarrow \nu_e) = 0.64 \pm 0.12$, calculated using the formula

$$P(\nu_e \rightarrow \nu_e) = \frac{R^{\text{exp}} - \Phi^{\text{SSM}} n_e \sigma_\mu}{\Phi^{\text{SSM}} n_e (\sigma_e + \sigma_\mu)}$$

where $R^{\text{exp}} = (167 \pm 19) \times 10^{-5}$ c.p.s. per 100 t is the experimental rate measured by Borexino, $\Phi^{\text{SSM}} = 5.98 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ is the pp neutrino theoretical flux, $n_e = (3.307 \pm 0.003) \times 10^{31}$ is the number of electrons for 100 t of the Borexino scintillator, and $\sigma_e = 11.38 \times 10^{-46} \text{ cm}^2$ and $\sigma_\mu = 3.22 \times 10^{-46} \text{ cm}^2$ are respectively the ν_e and ν_μ integrated cross-sections over the pp neutrino spectrum (which include radiative corrections^{25,26} (also P. Langacker and J. Erler, personal communication)).

All the other data points correspond to the previous Borexino results^{17–19}. We observe excellent agreement between our results and the MSW-LMA theoretical curve.

The standard solar model and the pp neutrino flux. The predictions for the solar neutrino fluxes come from solar models developed by astrophysicists since the 1960s. The SSM^{40,41} uses the simplest physical hypotheses and the best available physics input. It is assumed that energy is generated by nuclear reactions in the core of the star (see Extended Data Fig. 3 for the pp cycle) and is transported by radiation in the central part and by convection in the outer part. The basic evolution equation is the hydrostatic equilibrium between the outward radiative pressure

force and the inward gravitational force. One ingredient of the model, the metallicity Z/X (content of heavy elements relative to hydrogen), is under discussion at present (see, for example, ref. 9): the most recent determination of Z/X (AGS09) gives a lower value than the previous one (GS98), but the corresponding predictions significantly disagree with the helioseismology observations. The measurement of all solar neutrino components, which differ between solar models in these two metallicity scenarios, could help resolve the issue. The most recent calculations are those of ref. 9 (other models, such as the one described in ref. 24, give similar results). The predictions for the pp neutrino flux on Earth are $5.98 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ for the high-metallicity model and $6.03 \times (1 \pm 0.006) \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ for the low-metallicity model (there is a difference of less than 1%, and we use the flux for high metallicity throughout this paper). When combined with the latest values of the neutrino oscillation parameters²⁵, the SSM predicts 131 ± 2 c.p.d. per 100 t from pp neutrinos in Borexino. Neutrino oscillation parameters are key for translating the interaction rate measured in Borexino into a solar neutrino flux, as they provide the relative ratio between electron neutrinos and muon and tau neutrinos, which have different elastic scattering cross-section with electrons.

Event selection, position and energy reconstruction. The basic event selection cuts for solar neutrino analysis are (1) no coincidence with muon events³⁰ (a 300 ms veto is applied following muons crossing the scintillator and buffer volumes, and a 2 ms veto following muons crossing only the water tank) and (2) position reconstruction within the innermost volume of the detector (the fiducial volume). The second cut is necessary to eliminate background from radioactivity in the nylon vessels, the SSS and the PMTs; for this analysis, $R < 3.021$ m, $|z| < 1.67$ m (the geometrical centre of the sphere defines the origin of the coordinates x and y (in the horizontal plane) and z , as well as the origin from where, R , is measured).

A signal is recorded when more than a preset number of PMTs (25 until 15 February 2013, 20 thereafter) detect light within a 100 ns time window. Dedicated software then decides whether to classify this as a physics event or archive it as electronic noise or other instrumental effect. A ray-tracing algorithm triangulates photon arrival times at each struck PMT to determine the position of the event, which is assumed to be point-like.

The procedure to assign the correct energy to each event starts from the number of photoelectrons recorded by the PMTs. For β - and γ -events, the energy of an event is roughly proportional to the number of collected photoelectrons⁴². Notably, α -events display sizeable quenching of scintillation light, description of which is beyond the scope of this Article. The number of photons detected for a given type of event depends on its position, being maximal at the centre of the detector. We correct for this position dependence using a parameterized response obtained by calibrating the detector with known radioactive sources deployed at different positions. Source calibration data (mostly γ -ray emitters) are then used to tune Monte Carlo simulation software²⁸, allowing us to model the detector response correctly. These are in turn used to determine the energy response of the detector. We detect approximately 500 photoelectrons for a 1 MeV electron; this number progressively reduces as ageing, malfunctioning PMTs are taken offline.

A complete description of the data selection and analysis can be found in ref. 28, with some conditions relaxed specifically for this analysis. In particular, the energy threshold for acceptance of events used in this analysis was lowered with respect to previous analyses looking for higher-energy neutrino interactions. Also by contrast with what was done in previous measurements, no β -like selection condition (based on the different scintillation time profile of α - and β -particles) was used, as its relative efficiency for neutrino and background events could not be reliably evaluated at the lower energies characteristic of solar pp neutrino interactions.

Fit of the ^{14}C . The ^{14}C rate has been determined by looking at a sample of data in which the event causing the trigger is followed by a second event within the time acquisition window of 16 μs . This particular second-event selection bypasses threshold effects intrinsic to any self-triggering approach, clearly visible in Extended Data Fig. 4. The energy spectrum and its fit, using the theoretical ^{14}C β -emission shape²⁷, are shown in Extended Data Fig. 5. The ^{14}C spectrum fits the data well, yielding a ^{14}C rate of 40 ± 1 Bq per 100 t. The uncertainty accounts for systematic effects due to the stability of the detector response in time, knowledge of the ^{14}C spectral shape and fit conditions, including the fit energy range. We note that the relatively poorly known shape factor of the ^{14}C β -emission spectrum has a $<2\%$ effect on the pp neutrino result. The measured ^{14}C rate translates into a $^{14}\text{C}/^{12}\text{C}$ isotopic ratio of $(2.7 \pm 0.1) \times 10^{-18}$. The typical value for this ratio for atmospheric (that is, biologic) carbon is $\sim 10^{-12}$. ^{14}C is mainly produced by the nuclear reaction $^{14}\text{N}(n,p)^{14}\text{C}$ supported by cosmic rays in the upper atmosphere. The underground petroleum origin of the Borexino liquid scintillator explains its much smaller ^{14}C abundance.

Study of pile-up events. The pile-up component is determined using a data-driven method as follows. The real triggered events without any selection cuts are artificially overlapped with random data samples. The combined synthetic events are selected and reconstructed using the procedure applied to the regular data. By construction, the synthetic pile-up method accounts for all possible event pile-up

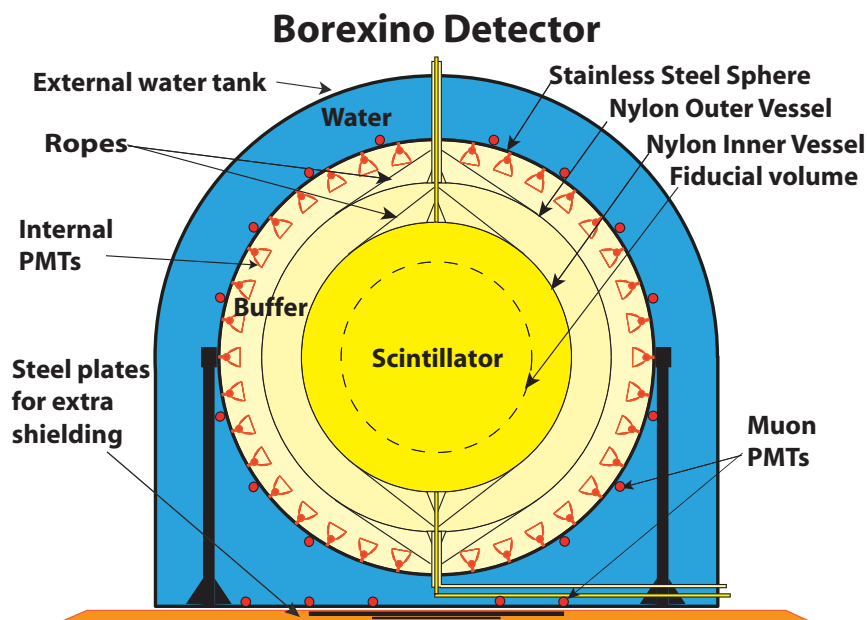
combinations. The corresponding spectrum includes events that vary in energy with respect to the original event (that is, before being artificially overlapped with the random sample) by more than a given number of photoelectrons (N_{\min}). In addition, the same event selection criteria as in the real data set are applied after the overlap. Thus, some systematic effects, for example the position reconstruction of pile-up events, are automatically taken into account. To increase the statistical precision, every real event is overlapped four times, each time with a different random data sample, and the final pile-up spectrum is divided by a factor of 4. An expanded view of the synthetic pile-up spectral probability density function (visible in Figs 2 and 3) is presented in Extended Data Fig. 6.

The final fit was performed using this synthetic pile-up with $N_{\min} = 5$ photoelectrons. This component is dominated but not exclusively composed of ^{14}C pile-up with itself. Some of the events are from dark noise of the PMTs. The robustness of the method was confirmed by checking that the fit results were not dependent on the choice of N_{\min} .

We compared the synthetic pile-up method with an alternative one. Regularly solicited trigger events with a $16\ \mu\text{s}$ acquisition time window (acquired at 0.5 Hz) are collected and sliced into fixed-time windows of the same duration as the signal window (230 and 400 ns). The hits in these windows produce the energy distribution shown in Extended Data Fig. 7, which represents a randomly sampled signal from the detector, including contributions from dark noise of the PMTs, ^{14}C and other radioactive contaminants. Pile-up can be thought of as the combination of events belonging to any spectral component combined with such a spectrum. The final fit can then be performed without a separate pile-up probability density function (as in the synthetic approach), by using 'smeared' spectral species in the final fit. The smearing is the convolution of ideal spectral components with the solicited trigger spectrum. Solar pp neutrino interaction rates measured using this method are in full agreement with those obtained with the synthetic pile-up method.

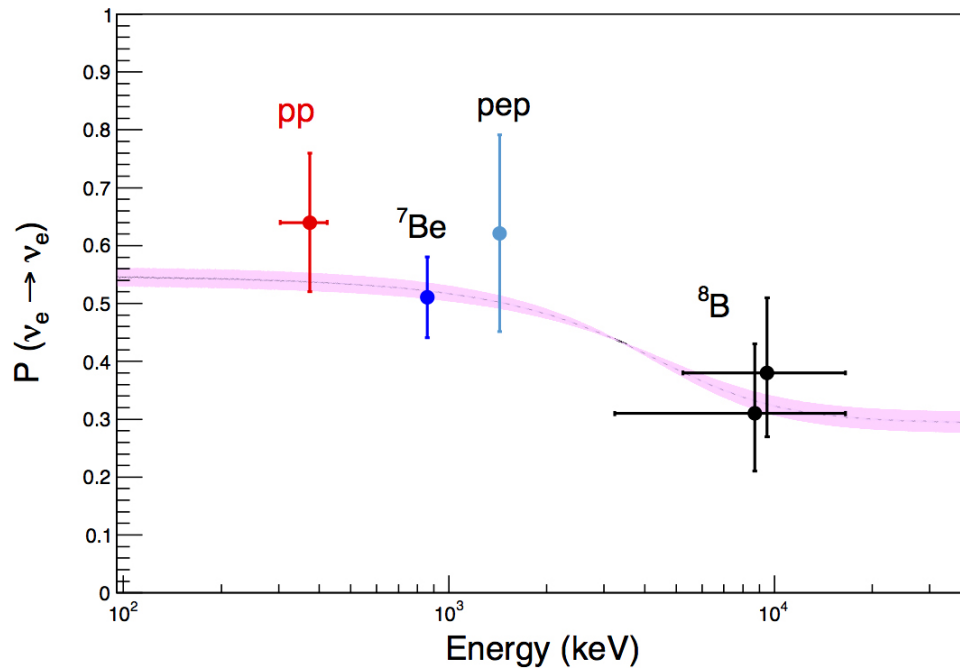
Stability of the result. The stability and robustness of the measured pp neutrino interaction rate was verified by performing fits with a wide range of different initial conditions, including fit energy range, synthetic-versus-convolution pile-up spectral shape, and energy estimator. The distribution of pp neutrino interaction rates obtained for all these fit conditions is summarized in Extended Data Fig. 8. The possibility of some remaining external background in the fiducial volume has been carefully studied in all Borexino solar neutrino analyses (see, for example, refs 17, 28). From these, we are confident that such a background at energies relevant for the pp neutrino study is negligible. In the particular case of the very low-energy part of the spectrum, we have tested this confidence by repeating the fit in five smaller fiducial volumes (with smaller radial and/or z -cut), which yields very similar results. Finally, the goodness (χ^2) of the spectral fit was computed using different values of the pp interaction rate, as shown in Extended Data Fig. 9. The absence of pp solar neutrinos is excluded with a statistical significance of 10σ .

36. Benziger, J. *et al.* A scintillator purification system for the Borexino solar neutrino detector. *Nucl. Instrum. Methods A* **587**, 277–291 (2008).
37. Alimonti, G. *et al.* The handling liquid systems for the Borexino solar neutrino detector. *Nucl. Instrum. Methods A* **609**, 58–78 (2009).
38. Alimonti, G. *et al.* (Borexino Collaboration). Ultra-low background measurements in a large volume underground detector. *Astropart. Phys.* **8**, 141–157 (1998).
39. Alimonti, G. *et al.* (Borexino Collaboration). Measurement of the ^{14}C abundance in a low-background liquid scintillator. *Phys. Lett. B* **422**, 349–358 (1998).
40. Bahcall, J. N. & Ulrich, R. K. Solar models, neutrino experiments and helioseismology. *Rev. Mod. Phys.* **60**, 297–372 (1988).
41. Bahcall, J. N. & Pinsonneault, M. H. Solar models with helium and heavy element diffusion. *Rev. Mod. Phys.* **67**, 781–808 (1995).
42. Knoll, G. F. *Radiation Detectors and Measurement* 4th edn, 290–291 (Wiley, 2010).



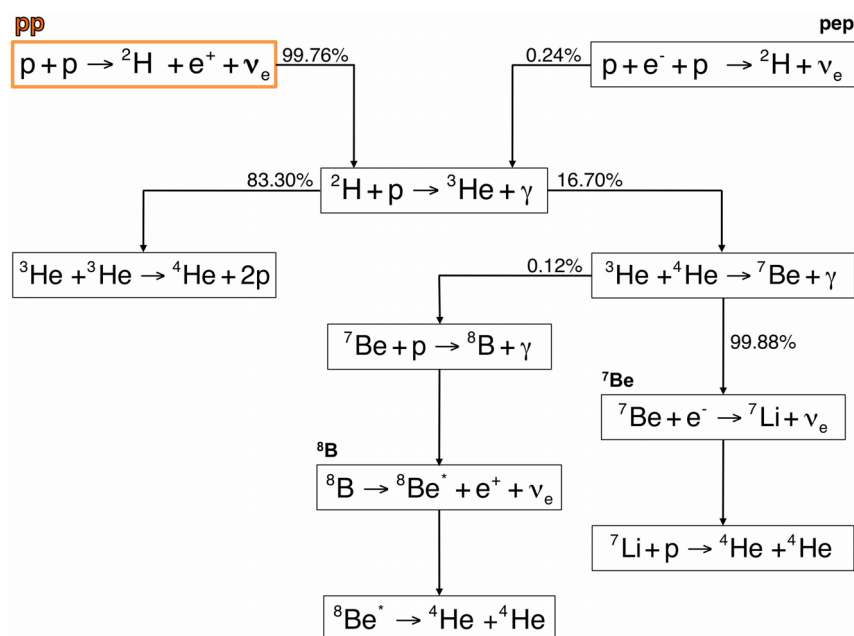
Extended Data Figure 1 | The Borexino detector. The characteristic onion-like structure of the detector²² is displayed, with fluid volumes of increasing radiological purity towards the centre of the detector. Although solar neutrino measurements are made using events whose positions fall inside the innermost volume of scintillator (the fiducial volume, shown as spherical for illustrative purposes only), the large mass surrounding it is necessary to

shield against environmental radioactivity. The water tank (17 m high) contains about 2,100 t of ultraclean water. The diameter of the stainless steel sphere is 13.7 m, and that of the thin nylon inner vessel containing the scintillator is 8.5 m. The buffer and target scintillator masses are 889 and 278 t, respectively.

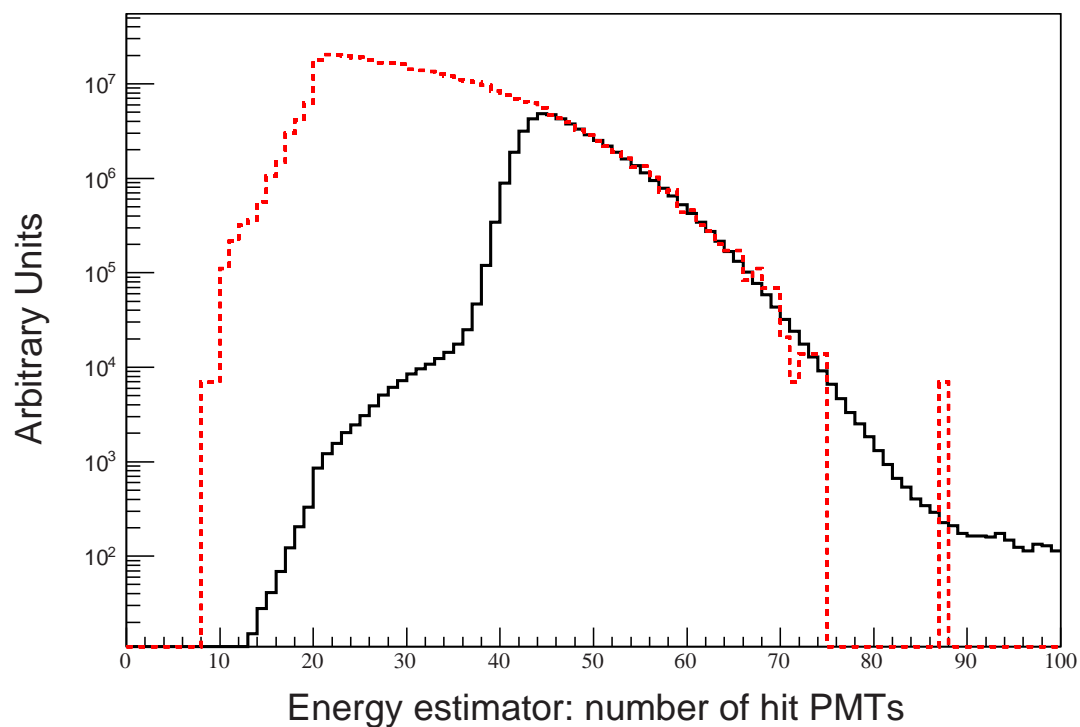


Extended Data Figure 2 | Survival probability of electron-neutrinos produced by the different nuclear reactions in the Sun. All the numbers are from Borexino (this paper for pp , ref. 17 for ${}^7\text{Be}$, ref. 18 for pep and ref. 19 for ${}^8\text{B}$ with two different thresholds at 3 and 5 MeV). ${}^7\text{Be}$ and pep neutrinos are mono-energetic. pp and ${}^8\text{B}$ are emitted with a continuum of energy, and the reported $P(\nu_e \rightarrow \nu_e)$ value refers to the energy range contributing to the

measurement. The violet band corresponds to the $\pm 1\sigma$ prediction of the MSW-LMA solution²⁵. It is calculated for the ${}^8\text{B}$ solar neutrinos, considering their production region in the Sun which represents the other components well. The vertical error bars of each data point represent the $\pm 1\sigma$ interval; the horizontal uncertainty shows the neutrino energy range used in the measurement.

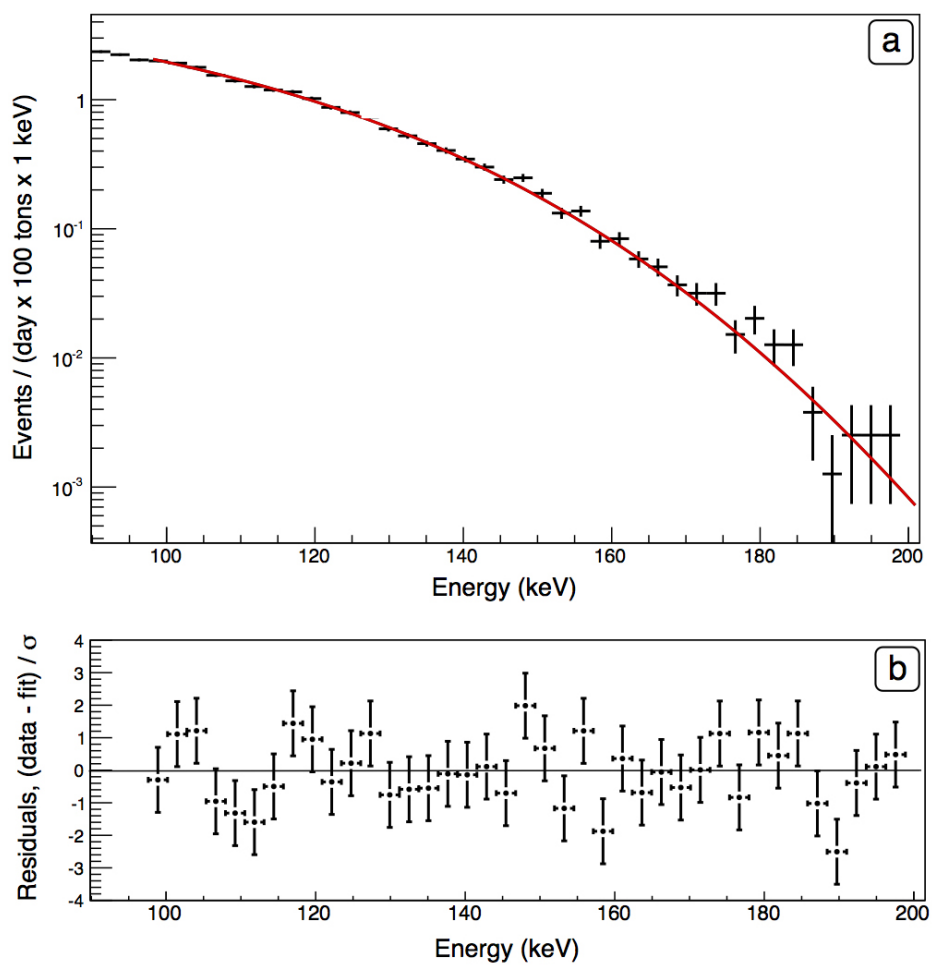


Extended Data Figure 3 | The sequence of nuclear fusion reactions defining the *pp* chain in the Sun. The *pp* neutrinos start the sequence 99.76% of the time.

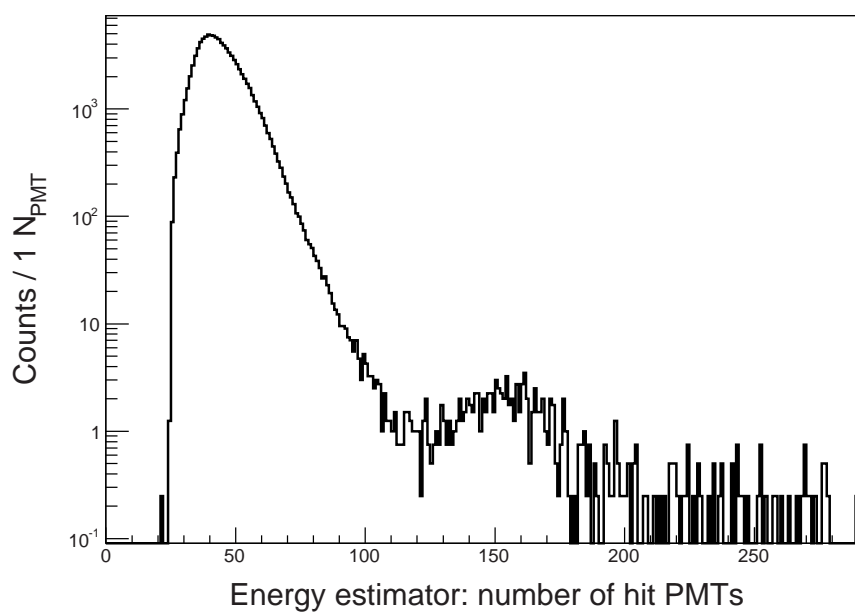


Extended Data Figure 4 | Study of the low energy part of the spectrum. Comparison of the spectrum obtained with the main trigger (black) and by selecting events falling in the late part of the acquisition window triggered by preceding events (red). Above 45 struck PMTs, the spectral shapes coincide.

The threshold effect for self-triggered events (black) is clear. The residual threshold effect at lower energy in the red curve is due to the finite efficiency for identifying very low-energy events within a triggered data window.

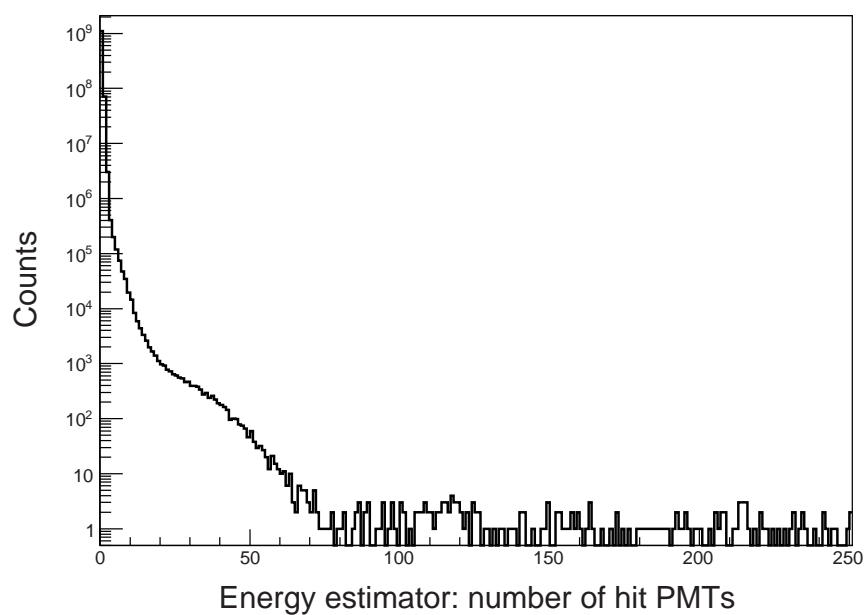


Extended Data Figure 5 | ^{14}C spectrum, and residuals, obtained from events triggered by a preceding event. **a**, Spectrum. **b**, Relative residuals of a fit with the ^{14}C β -emission spectrum (in units of standard deviations). The error bars thus represent $\pm 1\sigma$ intervals.

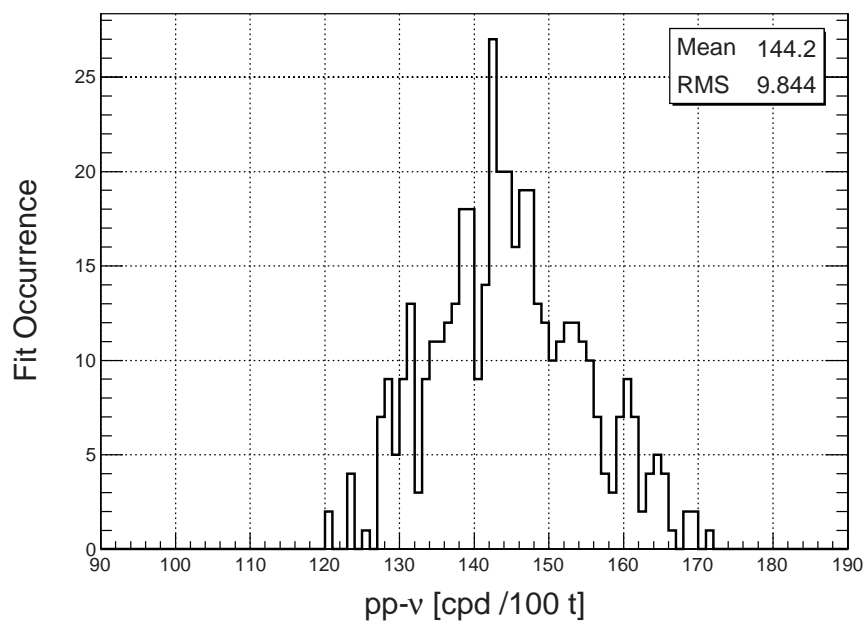


Extended Data Figure 6 | Energy spectrum of the pile-up data for the standard cuts. The small bump around 150 struck PMTs (~ 400 keV in

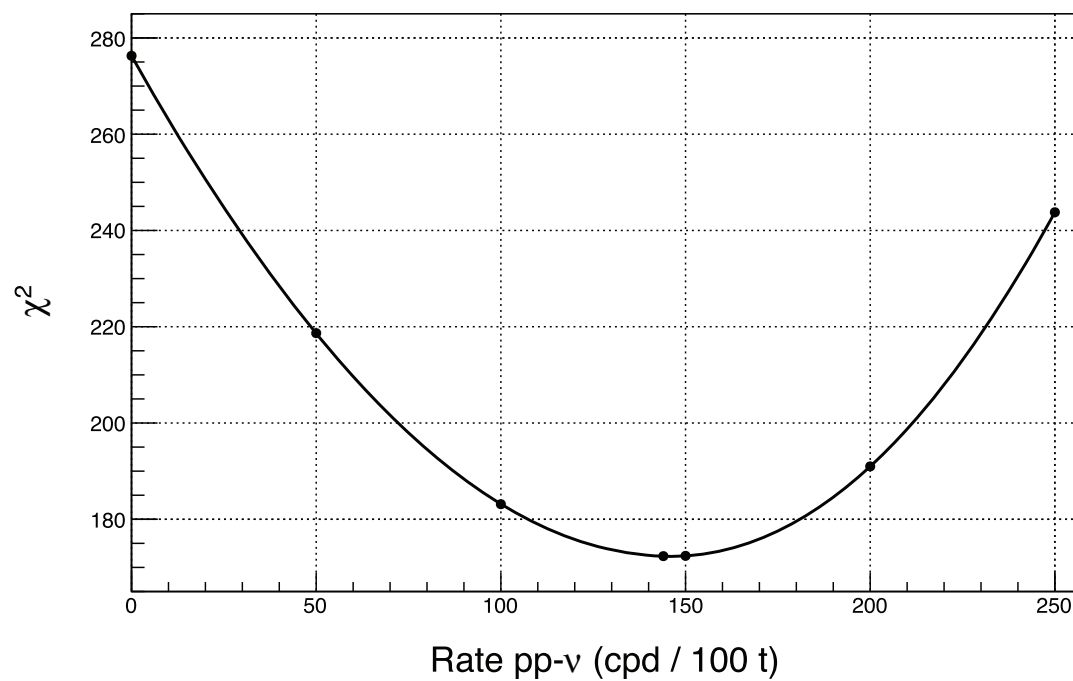
Figs 2 and 3) is due to the pile-up of ^{14}C with ^{210}Po ; at lower energies, pile-up is dominated by $^{14}\text{C}+^{14}\text{C}$, and by ^{14}C +dark noise.



Extended Data Figure 7 | Energy distribution of events collected with no threshold applied. The events correspond to regular, solicited triggers (sliced into 230 ns windows). This represents what the detector measures when randomly sampled. In an alternative treatment of pile-up, this spectrum is used to smear each spectral component used in the fit (see text for details).



Extended Data Figure 8 | Distribution of best-fit values for the pp neutrino interaction rate. Values are obtained by varying the fit conditions, including the fit energy range, synthetic-versus-analytic pile-up spectral shape, and energy estimator. The distribution shown is peaked around our reported value of 144 c.p.d. per 100 t.



Extended Data Figure 9 | Goodness of fit versus pp neutrino interaction rate. The χ^2 minimum is at our reported value of 144 c.p.d. per 100 t.

AhR sensing of bacterial pigments regulates antibacterial defence

Pedro Moura-Alves^{1*}, Kellen Faé^{1*}, Erica Houthuys^{1*}, Anca Dorhoi^{1*}, Annika Kreuchwig², Jens Furkert², Nicola Barison³, Anne Diehl², Antje Munder⁴, Patricia Constant⁵, Tatsiana Skrahina¹, Ute Gühlich-Bornhof¹, Marion Klemm¹, Anne-Britta Koehler¹, Silke Bandermann¹, Christian Goosmann⁶, Hans-Joachim Mollenkopf⁷, Robert Hurwitz⁸, Volker Brinkmann⁶, Simon Fillatreau⁹, Mamadou Daffe⁵, Burkhard Tümmler⁴, Michael Kolbe³, Hartmut Oschkinat², Gerd Krause² & Stefan H. E. Kaufmann¹

The aryl hydrocarbon receptor (AhR) is a highly conserved ligand-dependent transcription factor that senses environmental toxins and endogenous ligands, thereby inducing detoxifying enzymes and modulating immune cell differentiation and responses. We hypothesized that AhR evolved to sense not only environmental pollutants but also microbial insults. We characterized bacterial pigmented virulence factors, namely the phenazines from *Pseudomonas aeruginosa* and the naphthoquinone phthiocol from *Mycobacterium tuberculosis*, as ligands of AhR. Upon ligand binding, AhR activation leads to virulence factor degradation and regulated cytokine and chemokine production. The relevance of AhR to host defence is underlined by heightened susceptibility of AhR-deficient mice to both *P. aeruginosa* and *M. tuberculosis*. Thus, we demonstrate that AhR senses distinct bacterial virulence factors and controls antibacterial responses, supporting a previously unidentified role for AhR as an intracellular pattern recognition receptor, and identify bacterial pigments as a new class of pathogen-associated molecular patterns.

The aryl hydrocarbon receptor (AhR) is a ligand-dependent transcription factor that senses environmental toxins, notably halogenated and non-halogenated polycyclic aromatic hydrocarbons, such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)¹. Upon activation, AhR enters the nucleus and, together with aryl hydrocarbon receptor nuclear translocator (ARNT), binds dioxin-responsive elements. This leads to transcriptional activation of target genes including its own negative regulator, the AhR repressor (AHRN), and detoxifying monooxygenases, including *CYP1A1* and *CYP1B1* (refs 1, 2). These enzymes degrade some ligands to metabolites with decreased activity and increased water solubility, thus facilitating their removal³. Recent interest has shifted to endogenous AhR ligands, including tryptophan (Trp) degradation products, such as kynurenines and 6-formylindolo[3,2-b]carbazole (FICZ)⁴. Although downstream events of AhR activation by environmental toxins, including cell cycle control, apoptosis and haematopoiesis^{4,5}, have been intensively studied, the biological mechanisms regulated by natural AhR ligands are incompletely understood. Endogenous and dietary AhR ligands may play a role in immune control^{6,7}, for example, through transcriptional regulation of several pro-inflammatory cytokines^{8–10}. A role of AhR in T cell polarization^{11–14}, accumulation of innate lymphoid cells and protection against *Citrobacter rodentium* infection has been revealed recently^{7,15,16}.

We demonstrate that AhR (1) binds and senses distinct pathogen-associated molecular patterns (PAMPs); (2) regulates immune and degradation pathways in myeloid and epithelial cells in response to these PAMPs, and (3) plays a crucial role in host defence against bacterial pathogens producing AhR ligands. Thus, by integrating pathogen recognition and regulating effector functions (that is, detoxification as well as

induction of protective immunity), AhR plays a key role in innate defence against bacteria.

AhR senses pigmented virulence factors

We screened for AhR ligands of bacterial origin by molecular modelling. Pigmented virulence factors from pulmonary pathogens, that is, the phenazines from *P. aeruginosa* such as pyocyanin (Pyo), 1-hydroxyphenazine (1-HP), phenazine-1-carboxylic acid (PCA) and phenazine-1-carboxamide (PCN), as well as the naphthoquinone phthiocol (Pht) from *M. tuberculosis* were predicted to bind AhR. Their structures show features similar to TCDD (Fig. 1a and Extended Data Fig. 1a)¹⁷, and their low molecular weights allow them to diffuse and permeate cell membranes^{18,19}. *In vitro* studies revealed toxic effects of reactive oxygen intermediates generated by Pht²⁰, similar to *P. aeruginosa* phenazines²¹. Additionally, phenazines exert various effects on host cells (for example, direct damage, cytokine production, inhibition of ciliary motion and mucus production)^{22–25} and their abundance negatively correlates with lung function in cystic fibrosis patients²⁶.

As the three-dimensional structure of AhR remains unknown, its interactions with these molecules were interrogated by *in silico* docking studies. The hydrophobic AhR binding pocket contains proton donors, whereas the pigments show two proton acceptors and fit well into the binding pocket (Fig. 1b, Extended Data Fig. 1b and Supplementary Information). In subsequent experiments we focused on Pyo, 1-HP and Pht.

To verify our *in silico* predictions, we evaluated ligand binding to AhR by intrinsic fluorescence quenching or a standard competition assay

¹Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, 10117 Berlin, Germany. ²Leibniz Institute for Molecular Pharmacology (FMP), Robert-Rössle-Strasse 10, 13125 Berlin, Germany. ³Max Planck Institute for Infection Biology, Structural Systems Biology, Charitéplatz 1, 10117 Berlin, Germany. ⁴Clinical Research Group, Clinic for Pediatric Pneumology, Allergology and Neonatology, OE 6710, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. ⁵Institute of Pharmacology and Structural Biology (IPBS), CNRS and University of Toulouse (Toulouse III), 205 Route de Narbonne, 31077, Toulouse cedex 04, Toulouse, France. ⁶Microscopy Core Facility, Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, 10117 Berlin, Germany. ⁷Microarray Core Facility, Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, 10117 Berlin, Germany. ⁸Protein Purification Core Facility, Max Planck Institute for Infection Biology, Charitéplatz 1, 10117 Berlin, Germany. ⁹German Rheumatism Research Centre Berlin (DRFZ), a Leibniz Institute, Charitéplatz 1, 10117 Berlin, Germany.

*These authors contributed equally to this work.

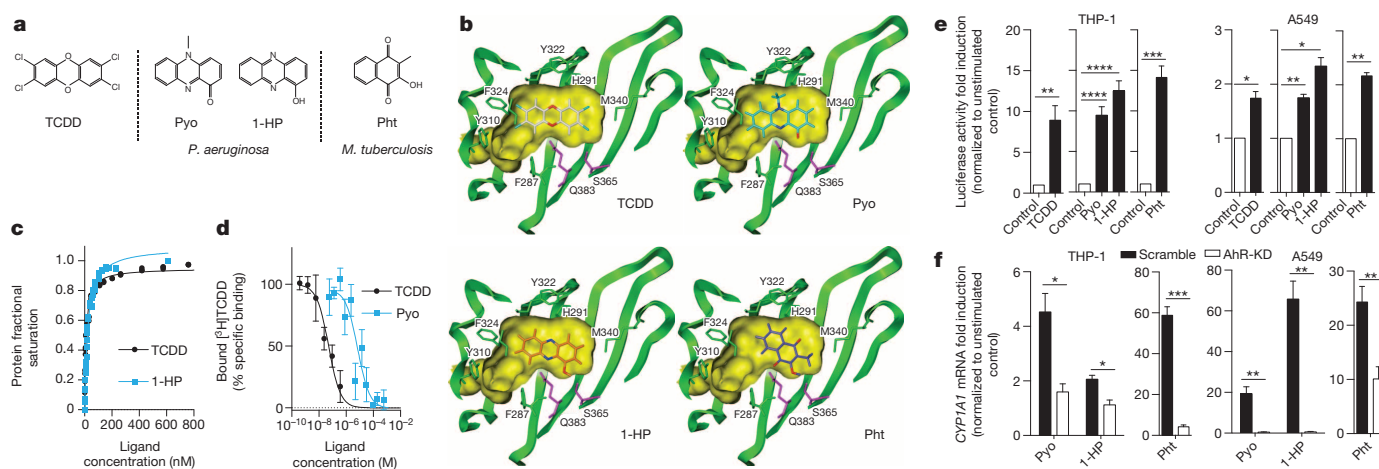


Figure 1 | Bacterial pigmented virulence factors activate AhR. **a**, Structures of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), pyocyanin (Pyo), 1-hydroxyphenazine (1-HP) and phthiocol (Pht). **b**, *In silico* docking into the AhR ligand-binding pocket (yellow). Hydrophilic residues (magenta), aromatic and hydrophobic residues (green). **c**, Intrinsic tryptophan fluorescence quenching of

using the radioactive ligand TCDD⁶. In the first case, a purified construct comprising only the major domains was used (amino acids 1 to 417, AhR₁₋₄₁₇), in the second the intact full-length receptor. Hyperbolic fitting of AhR₁₋₄₁₇ fluorescence changes titrated with TCDD and 1-HP yielded dissociation constant (K_d) values of 11.1 and 23.5 nM, respectively (Fig. 1c and Extended Data Fig. 1c). Due to the fluorescence overlap of Pyo and Trp, binding of Pyo to AhR was not studied by ultraviolet but by concentration-dependent displacement of [³H]TCDD from wild-type (WT) mouse liver cytosol extracts (Fig. 1d)⁶. TCDD showed an inhibition constant K_i = 40 nM (95% confidence limits: 15–107 nM), whereas the K_i of Pyo was 5.4 μ M (95% confidence limits: 2.0–15.1 μ M). Binding of Pht to AhR could not be measured with this assay due to its sensitivity towards de-oxidation by dithiothreitol required for stable assay conditions.

Ligand-induced AhR activation was verified by a luciferase reporter assay with human macrophages (THP-1), type II pneumocytes (A549) and bronchiolar Clara (H358) cell lines (Fig. 1e and Extended Data Fig. 1d, e). All pigments activated AhR in a dose-dependent manner (Extended Data Fig. 1f). Concentrations of up to 100 μ M of Pyo have been detected in pulmonary secretions of cystic fibrosis patients infected with *P. aeruginosa*^{26,27}. Pht concentrations of 50 μ M have been estimated to be present in *M. tuberculosis*²⁰. Thus bacterial ligands were used at concentrations of 50 μ M in further experiments, which did not affect cell viability (Extended Data Fig. 1g). AhR was not activated upon stimulation of THP-1 AhR reporter cells with known Toll-like receptor (TLR) ligands (Extended Data Fig. 1h). Bacterial pigments significantly induced transcription of hallmark genes of AhR activation, including *CYP1A1*, *CYP1B1* and *AHR* (Extended Data Fig. 2a). Silencing of AhR (knockdown KD, AhR-KD cells; Extended Data Fig. 2b) decreased *CYP1A1* induction after ligand stimulation (Fig. 1f and Extended Data Fig. 2c). Consistently, microarray analysis of A549 cells stimulated with bacterial pigments revealed differential regulation of the AhR pathway (Extended Data Fig. 3a, b and Supplementary Tables 1 and 2). AhR can be activated indirectly via inhibition of *CYP1A1* and increased abundance of Trp metabolite FICZ²⁸. However, AhR activation was still observed in absence of Trp in the medium (Extended Data Fig. 4a) and *AHR* was induced upon stimulation in both Scramble and *CYP1A1*-KD THP-1 cells (Extended Data Fig. 4b–d). Hence, AhR was activated by these pigments independently of *CYP1A1* inhibition. In summary, AhR directly senses pigmented bacterial virulence factors thereby mobilizing the detoxification pathway.

AhR–*P. aeruginosa* crosstalk via phenazines

Infection of the THP-1 AhR reporter cell line with the laboratory strain *P. aeruginosa* PAO1 or stimulation with filtered supernatants from

purified AhR₁₋₄₁₇. **d**, Displacement of [³H]TCDD from AhR in WT mouse liver cytosol. **e**, AhR reporter cells stimulated for 24 h. **f**, qRT-PCR of *CYP1A1* at 4 h (THP-1) and 24 h (A549) after stimulation. **c**, **d**, Results of two experiments; **e**, **f**, cumulative data of at least two experiments, mean + s.e.m.; **e**, **f**, Student's *t*-test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

PAO1 cultures activated AhR (Fig. 2a, b). Mutants of the *P. aeruginosa* PA14 strain overexpressing Pyo and/or all phenazines were consistently stronger activators than the parental WT strain (PA14-WT1) (Fig. 2c, d and Extended Data Fig. 5a–c and Supplementary Table 3). Supernatants of the PA14 Δ phz1/2 strain, which lack the production of all phenazines²⁹, induced significantly lower AhR activation compared to supernatants of the parental WT (PA14 WT2) (Fig. 2e, f and Extended Data 5a,d). These data define phenazines as major AhR activators in *P. aeruginosa*.

Degradation of Pyo has been described as a host defence stratagem against *P. aeruginosa* although the underlying mechanisms remain controversial^{30–32}. Pyo degradation was observed in A549 cells and knockdown of AhR reduced Pyo degradation, as determined by spectrophotometry (Fig. 2g) and HPLC analysis (Fig. 2h and Extended Data Fig. 6a). Similar results were obtained for *M. tuberculosis* Pht (Extended Data Fig. 6b). Thus, sensing of bacterial pigments by AhR caused their direct degradation. AhR regulates the expression of several xenobiotic metabolizing enzymes (XME)^{2,33}. Microarray analysis of A549 cells stimulated with bacterial pigments identified Phase I XMEs (*CYP1A1* and *CYP1B1*) amongst the strongest upregulated genes in Scramble as compared to AhR-KD cells (Extended Data Fig. 6c and Supplementary Tables 4 and 5). Although direct detoxification of environmental toxins is a well-described mechanism, its relevance in immune defence against pathogens has been largely ignored.

AhR in *P. aeruginosa* infection

We used *Ahr*^{−/−} mice to determine AhR contribution to host defence against pulmonary infection with the opportunistic pathogen *P. aeruginosa*^{21,34–36}. *Ahr*^{−/−} mice developed severe disease and respiratory distress after infection with PAO1 (Fig. 3a and Extended Data Fig. 7a, b). Lung bacterial loads and tissue damage were consistently exacerbated in *Ahr*^{−/−} mice (Fig. 3b, c and Extended Data Fig. 7c) which succumbed to *P. aeruginosa* infection more rapidly than WT mice (Fig. 3d). Absence of AhR fostered alveolar disruption, vascular leakage and deposition of bacterial microcolonies in the alveoli (Fig. 3e). Neutrophils are essential for early defence against *P. aeruginosa*³⁴. Their numbers were significantly reduced in lung tissue and bronchoalveolar lavage fluid (BALF) of *Ahr*^{−/−} mice promptly post-infection (p.i.) (Fig. 3f and Extended Data Fig. 8a, b), consistent with reduced cell infiltration (Fig. 3e). Abundance of myeloid cells in uninfected *Ahr*^{−/−} versus WT mice and lymphoid cell frequencies in infected *Ahr*^{−/−} versus WT mice did not differ significantly (Extended Data Fig. 8c–e). Frequencies of necrotic and apoptotic leukocytes in BALF from WT and *Ahr*^{−/−} mice were comparable, suggesting defective recruitment rather than increased cell death in infected *Ahr*^{−/−} mice (Fig. 3g). Accordingly, concentrations of neutrophil chemoattractants³⁷, CXCL1

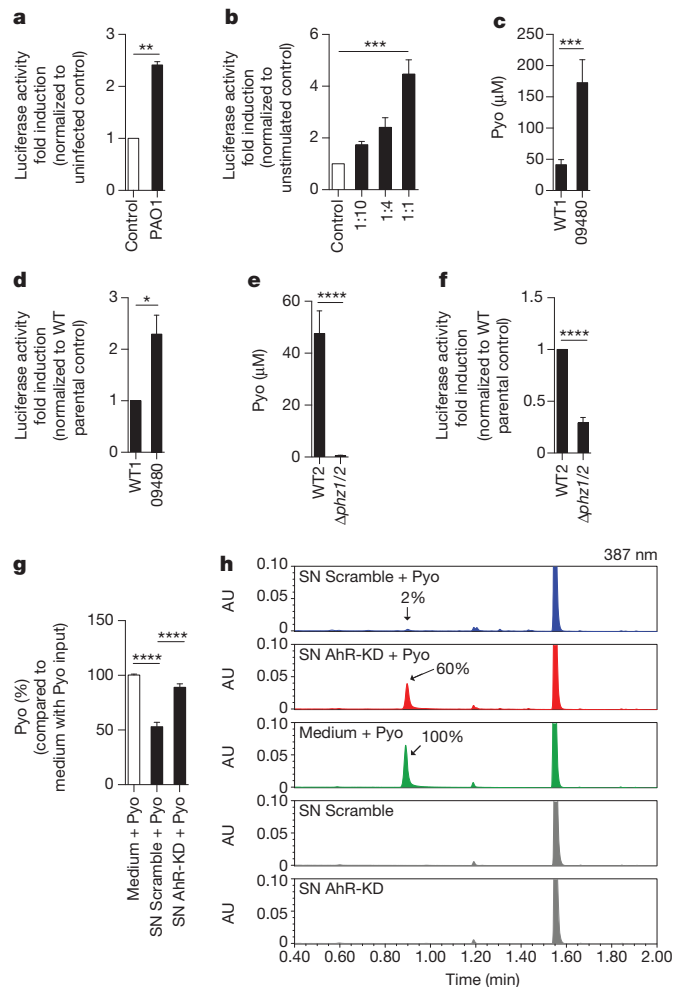


Figure 2 | AhR activated by secreted *P. aeruginosa* compounds induces pyocyanin degradation. **a, b, d, f,** AhR reporter THP-1 cells after 24 h (**a**) infection with *P. aeruginosa* PAO1 or **b, d, f,** stimulation with culture supernatants of: (**b**) PAO1, (**d, f**) PA14 parental wild-type (WT, PA14 WT1, PA14 WT2) and mutant (PA14 09480, PA14 Δphz1/2) strains. **c, e,** Pyo concentrations in culture supernatants of PA14 strains. **g, h,** Spectrophotometry (**g**) and HPLC (**h**) of Pyo degradation in supernatants (SN) of A549 cells 48 h after Pyo stimulation. Arrows depict Pyo at 0.9 min. **a–g,** Cumulative data of at least three experiments, mean + s.e.m.; **h,** representative of three experiments; **a, c–f,** Student's *t*-test; **b, g,** one-way ANOVA. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

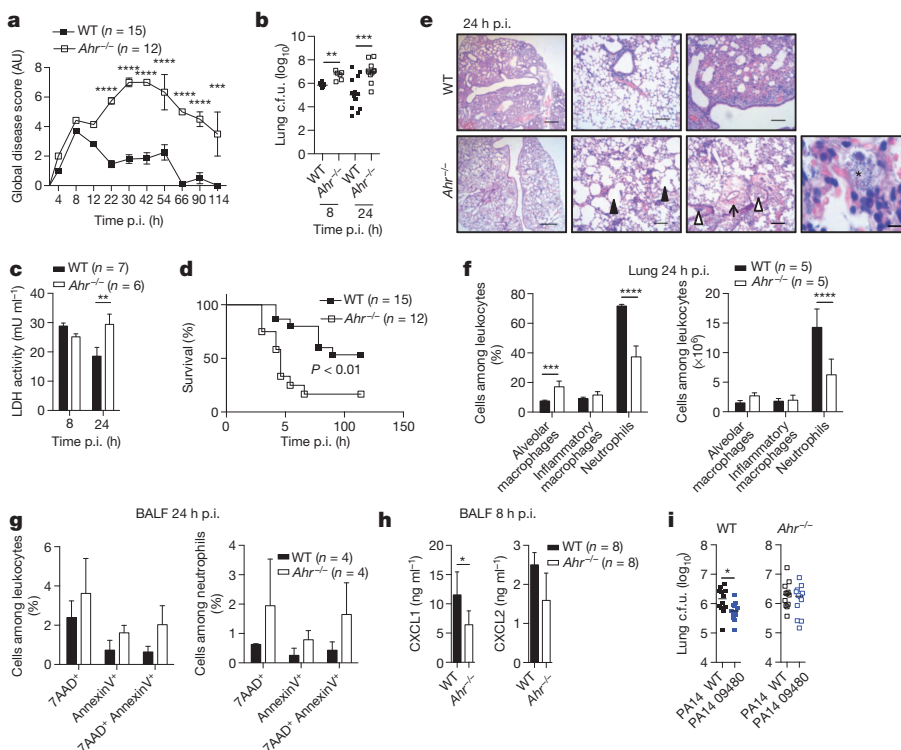
and CXCL2, were reduced in BALF from infected *Ahr*^{−/−} mice (Fig. 3h). No differences were observed for most other pro-inflammatory cytokines and chemokines analysed (Extended Data Fig. 8f). We therefore attribute reduced survival of *Ahr*^{−/−} mice primarily to impaired neutrophil recruitment early post-infection leading to higher bacterial burden and exacerbated tissue damage.

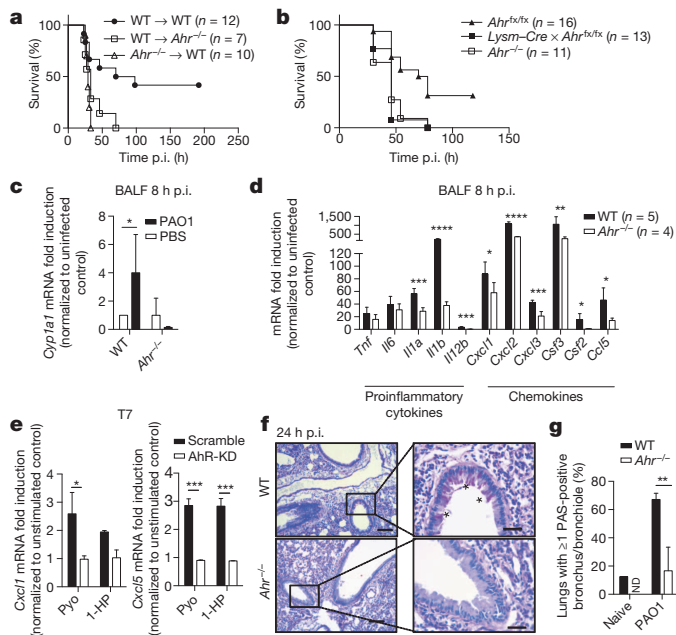
To interrogate the role of AhR in sensing phenazines *in vivo*, we infected mice with *P. aeruginosa* PA14 09480, which expresses higher amounts of all phenazines as compared to its parental WT (PA14 WT1, Extended Data Fig. 5b). The mutant was cleared faster from the lungs of WT mice (Fig. 3i). In contrast, increased resistance to the mutant was abolished in *Ahr*^{−/−} mice, revealing AhR sensing of these ligands *in vivo*. Thus, upon infection with *P. aeruginosa*, early sensing of bacterial phenazines by AhR reduced bacterial burden in the lung favouring host resistance.

To identify the cell type responsible for AhR sensing of phenazines, we infected bone marrow chimaeras with PAO1. Susceptibility to infection was comparable in WT→*Ahr*^{−/−} and *Ahr*^{−/−}→WT (arrow indicates the transfer of cells in chimaeric mice) chimaeras (Fig. 4a), suggesting that AhR signalling in radiosensitive haematopoietic and radioresistant cells is non-redundant for AhR-induced antibacterial defence. Infection of myeloid-specific AhR-deficient mice (*Lysm*−*Cre* × *Ahr*^{flx/flx}) (Fig. 4b) revealed that among haematopoietic cells, AhR-regulated factors in macrophages and neutrophils play a critical role in protection against *P. aeruginosa*.

Figure 3 | AhR regulates host defence against *P. aeruginosa*.

WT and AhR-deficient (*Ahr*^{−/−}) mice were infected intratracheally with 4×10^6 (**b, c, e–h**) or 10^7 colony-forming units (c.f.u.) (**a, d**) of *P. aeruginosa* PAO1 or 2×10^6 c.f.u. of PA14 (WT1 or 09480) (**i**). **a,** Global disease score of surviving animals. **b, i,** Bacterial burdens in lungs. **c,** Lactate dehydrogenase (LDH) activity in serum. **d,** Mouse survival. **e,** Haematoxylin and eosin staining of lung. Disrupted and dilated alveoli (black arrowheads), tissue debris in bronchiole (white arrowheads), fibrin deposits (arrow) and bacteria in alveolus (asterisks). Scale bars, 400 μm (left), 100 μm (middle), 10 μm (lower right). **f, g,** FACS analysis of leukocytes. **h, CXCL1** and **CXCL2** concentrations. **a–d, h, i,** Cumulative data of two independent experiments; **e–g,** representative of two experiments; **a, c,** two-way ANOVA, mean + s.e.m.; **b, i,** Mann–Whitney *U*-test, median; **d,** Log-rank test; **f, g,** two-way ANOVA, mean + s.d.; **h, i,** Mann–Whitney *U*-test, median + interquartile range. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.





The canonical AhR pathway was activated in leukocytes (similar proportions of neutrophils and inflammatory macrophages (Extended Data Fig. 8b)), derived from BALF of *P. aeruginosa*-infected mice (Fig. 4c and Extended Data Fig. 8g, h). Microarray analysis unveiled AhR-dependent differentially regulated genes and functions (Supplementary Tables 6–9). AhR-regulated pro-inflammatory cytokine and chemokine genes were

Figure 4 | Haematopoietic and non-haematopoietic cells contribute to AhR-mediated defence against *P. aeruginosa*. **a**, Bone marrow chimaeras and **b**, *Lysm-Cre* × *Ahr*^{fx/fx}, WT controls (*Ahr*^{fx/fx}) and *Ahr*^{-/-} mice were infected with 10⁷ c.f.u. of *P. aeruginosa* PAO1. **c**, **d**, qRT-PCR analysis of leukocytes. **e**, qRT-PCR of *Cxcl1* and *Cxcl5* in pneumocytes stimulated for 4 h. **f**, **g**, Periodic Acid Schiff (PAS) staining of lungs; PAS⁺ cells are indicated by asterisks. Scale bars, 100 μm (left), 30 μm (right). **a**, One experiment; **b**, cumulative data of two experiments; **c**, **d**, representative of two experiments, mean + s.d.; **e**, cumulative data of three experiments, mean + s.e.m.; **g**, cumulative data of two experiments (naive WT *n* = 8, *Ahr*^{-/-} *n* = 5; PAO1 WT *n* = 12, *Ahr*^{-/-} *n* = 12), mean + s.e.m.; **a**, **b**, Log-rank test; **c**–**e**, Student's *t*-test; **g**, two-way ANOVA. **a**, WT → *Ahr*^{-/-} vs WT → WT *P* < 0.05; *Ahr*^{-/-} → WT vs WT → WT *P* < 0.01; *Ahr*^{-/-} → WT vs WT → *Ahr*^{-/-} NS; **b**, *Lysm-Cre* × *Ahr*^{fx/fx} vs *Ahr*^{fx/fx} *P* < 0.01; *Lysm-Cre* × *Ahr*^{fx/fx} vs *Ahr*^{-/-} NS; *Ahr*^{fx/fx} vs *Ahr*^{-/-} *P* < 0.01. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. ND, not detected; NS, not significant.

confirmed by qRT-PCR (Fig. 4d). A crosstalk between AhR and other signalling pathways has been described^{38,39}, probably accounting for a more complex regulation of these genes.

Respiratory epithelial cells are well positioned for prompt initiation of antibacterial defence through cytokines and chemokines⁴⁰. Induction of neutrophil chemoattractants in mouse type II pneumocytes (T7 cell line) by Pyo/1-HP depended on AhR expression (Fig. 4e and Extended Data Fig. 2b). Epithelial cells also serve as primary barrier against pathogen invasion by producing mucus and surfactants⁴⁰. A role for AhR in mucus production in response to TCDD has been addressed⁴¹. *P. aeruginosa* induced mucus production in bronchiolar and bronchial cells of WT mice, whereas mucus-positive cells were hardly detected in *Ahr*^{-/-} mice (Fig. 4f, g). We conclude that AhR sensing of bacterial pigments promptly triggers natural and innate defence mechanisms in myeloid and parenchymal cells, as exemplified by induction of inflammatory

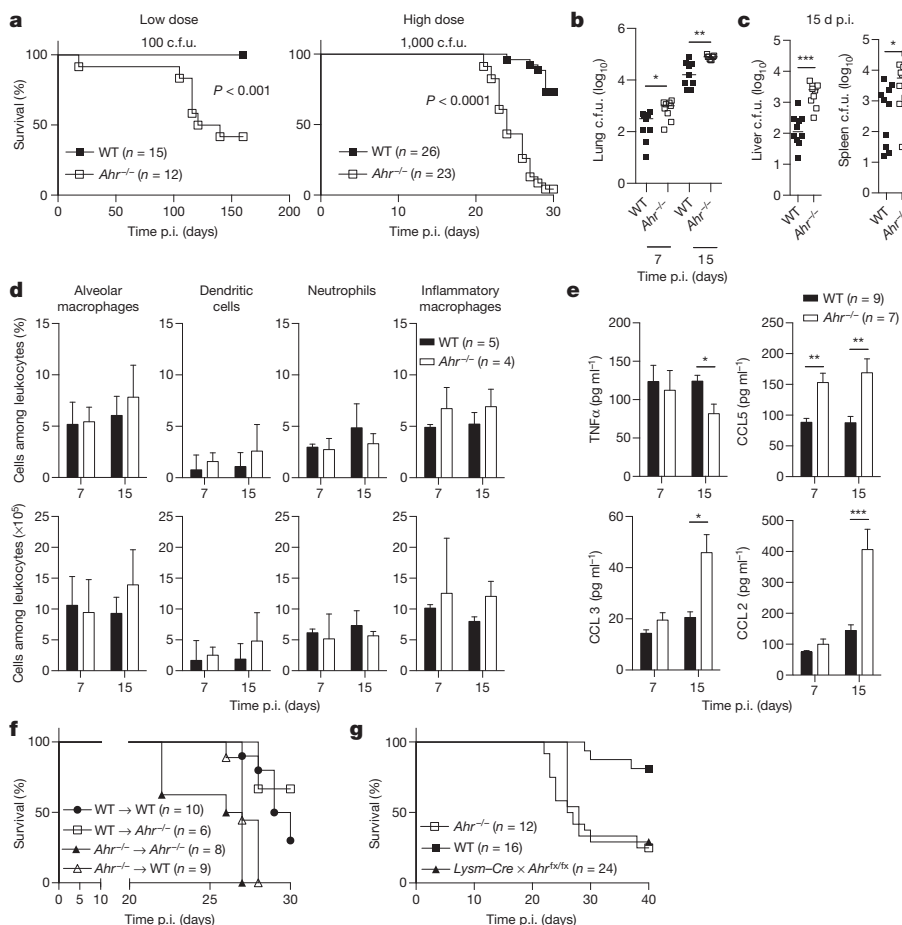


Figure 5 | AhR is critical for host defence against *M. tuberculosis*. **a**, WT and *Ahr*^{-/-} mice were infected with low-dose (**a** (left), **b**–**e**) or high-dose (**a** (right), **f**, **g**) *M. tuberculosis* H37Rv. **b**, **c**, Bacterial burdens. **d**, FACS analysis of pulmonary leukocytes. **e**, Cytokines and chemokines in lung homogenates. **f**, **g**, Mouse survival. **a**–**c** Cumulative data of (**a** (left), **b**, **c**) two or (**a** (right)) four experiments; **d**, representative of two experiments, mean + s.d.; **e**, cumulative data of two experiments, mean + s.e.m.; **f**, representative of two experiments; **g**, cumulative data of two experiments; **a**, Log-rank test; **b**, **c**, Mann–Whitney *U*-test, median; **d**, two-way ANOVA; **e**, Mann–Whitney *U*-test; **f**, WT → WT vs *Ahr*^{-/-} → *Ahr*^{-/-} *P* < 0.0001; *Ahr*^{-/-} → WT vs WT → WT *P* < 0.001; WT → *Ahr*^{-/-} vs WT → WT NS; WT → *Ahr*^{-/-} vs *Ahr*^{-/-} → WT *P* < 0.01; WT → *Ahr*^{-/-} vs *Ahr*^{-/-} → *Ahr*^{-/-} *P* < 0.001; *Ahr*^{-/-} → WT vs *Ahr*^{-/-} → *Ahr*^{-/-} *P* < 0.05. **g**, *Lysm-Cre* × *Ahr*^{fx/fx} vs *Ahr*^{-/-} NS; *Lysm-Cre* × *Ahr*^{fx/fx} vs WT *P* < 0.001, WT vs *Ahr*^{-/-} *P* < 0.001. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

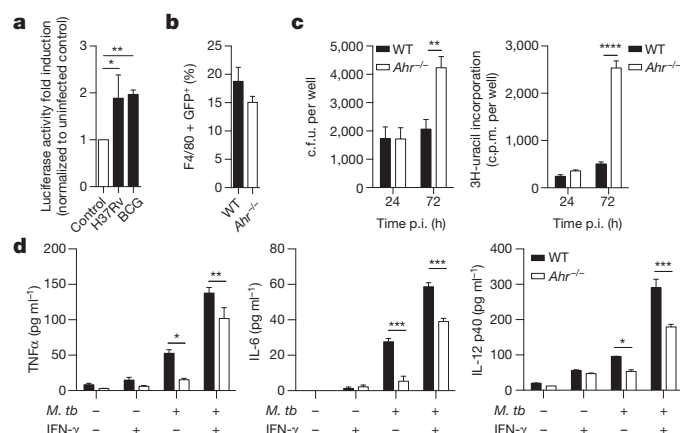


Figure 6 | AhR mediates macrophage activation and mycobacterial growth inhibition. **a**, AhR reporter-infected THP-1 cells. **b–d**, Bone-marrow-derived macrophages (BMDMs) were infected with *M. tuberculosis* (multiplicity of infection (m.o.i.) = 5). **b**, Uptake of H37Rv-GFP at 4 h p.i. **c**, Bacterial growth. **d**, Cytokines in supernatants at 24 h p.i. **a–d** Cumulative data of at least two experiments, mean \pm s.e.m.; **a**, one-way ANOVA; **b**, Mann–Whitney U-test; **c**, **d**, two-way ANOVA. * P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001.

cytokines, neutrophil recruitment to, and mucus production at, the site of infection.

AhR in *M. tuberculosis* infection

To corroborate the generality of AhR sensing in pulmonary infections, we determined its contribution to host defence against the persistent intracellular pathogen *M. tuberculosis* *in vivo*. *Ahr*^{−/−} mice succumbed to low-dose aerosol infection, but only at late time points (≥ 100 days) (Fig. 5a). Yet, by 7 days post-infection, bacterial load was significantly elevated in lung, liver and spleen of *Ahr*^{−/−} mice (Fig. 5b, c), pointing to a critical role of AhR in control of innate pulmonary defence and in preventing dissemination of *M. tuberculosis*. No significant differences in tissue damage were observed (Extended Data Fig. 9a, b), although a trend towards elevated numbers of inflammatory macrophages was observed in lungs of *Ahr*^{−/−} versus WT mice at day 7 post-infection (Fig. 5d). Abundance of TNF- α was decreased (Fig. 5e), and macrophage chemoattractants CCL2, CCL3 and CCL5 were increased in *Ahr*^{−/−} lung homogenates (Fig. 5e), correlating with the kinetics of the mononuclear phagocytes. Lymphocyte numbers (Extended Data Fig. 9c), and concentrations of most other cytokines and chemokines did not differ (Extended Data Fig. 9d).

Coherent with a decisive role for AhR in early tuberculosis, after high-dose aerosol *M. tuberculosis* infection, mortality of *Ahr*^{−/−} versus WT mice was exacerbated (Fig. 5a). In infected bone marrow chimerae, resistance was increased in all experimental groups over *Ahr*^{−/−} \rightarrow *Ahr*^{−/−} mice (Fig. 5f) and no statistically significant differences were observed between WT \rightarrow WT versus WT \rightarrow *Ahr*^{−/−} (Fig. 5f). Ameliorated survival of *Ahr*^{−/−} \rightarrow WT over *Ahr*^{−/−} \rightarrow *Ahr*^{−/−} mice point to an additional role of AhR sensing by non-haematopoietic cells. Consistent with a central role of AhR in the hematopoietic compartment, survival did not differ significantly between *Lysm*-Cre \times *Ahr*^{flx/flx} versus *Ahr*^{−/−} mice (Fig. 5g). The critical role of myeloid cells in AhR-dependent host defence against tuberculosis is congruent with a critical function of macrophages in this infection⁴².

AhR-dependent luciferase activity was induced in *M. tuberculosis* and *M. bovis* BCG-infected macrophage reporter cells (Fig. 6a). Furthermore, AhR-regulated genes were upregulated upon infection of primary human macrophages with *M. tuberculosis* (Extended Data Fig. 9e). Internalization of *M. tuberculosis* by *Ahr*^{−/−} bone-marrow-derived macrophages (BMDMs) remained unaffected (Fig. 6b), but mycobacterial growth inhibition and secretion of the pro-inflammatory cytokines TNF- α , IL-12p40, and IL-6 were impaired in *Ahr*^{−/−} BMDMs (Fig. 6c, d).

Intriguingly, Pht was detected in lipid fractions of *M. tuberculosis*, which activated AhR, while non-activating lipid fractions were devoid of Pht (Extended Data Fig. 10). Although we are confident that Pht sensing by AhR is critical for prompt mobilization of host defence in tuberculosis, we cannot exclude the existence of other mycobacterial or endogenous ligands, notably at later stages of tuberculosis⁴³.

Discussion

Our findings extend the list of AhR ligands beyond environmental toxins, endogenous molecules and dietary components to bacterial virulence factors. We demonstrate that: (1) AhR binds bacterial pigments comprising a redox-cycling phenazine/naphthoquinone moiety, namely, *P. aeruginosa* Pyo, 1-HP, and probably PCA and PCN, as well as *M. tuberculosis* Pht; (2) AhR sensing leads to transcription of canonical detoxifying genes as well as regulation of cytokine and chemokine production; (3) AhR activation regulates inflammatory leukocyte recruitment to the infected lung and control of bacterial replication; and (4) AhR plays a central role in defence against both acute and chronic bacterial infections.

The family of phenazines represents an immense class of secondary metabolites present only in prokaryotes⁴⁴ but not restricted to *Pseudomonas* spp.^{19,44,45}. Naphthoquinones are broadly distributed in prokaryotes, plants and animals²⁰. As a corollary, we speculate that the broad distribution of these pigments has served as a driving force for AhR evolution in host defence against exogenous insult. Concepts of how the innate immune system activates host defence against bacteria have focused on direct attack of invading pathogens, followed by instruction of acquired immunity^{46,47}. In contrast, maintenance of host integrity threatened by environmental toxins has emphasized rapid detoxification. Our data draw attention to the need for a conceptual convergence of both types of responses for pathogens producing virulence factors, which directly affect host cells. The AhR is well equipped for both functions by concomitantly mobilizing virulence factor neutralization and bacterial eradication. Sensing of PAMPs by pattern recognition receptors (PRRs) expressed by immune cells, notably TLR cognates, alerts the host of invading pathogens^{46,47}. Here, we propose bacterial pigments as a new class of PAMPs, and as a corollary, AhR as a new PRR. Defining AhR as a PRR, which senses PAMPs (phenazines/naphthoquinones as shown here) and self-components associated with inflammation—also termed danger-associated molecular patterns (DAMPs)—indicates biological similarities rather than disparities between the two types of surveillance mechanisms in control of bacterial infections.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 October 2012; accepted 17 July 2014.

Published online 13 August 2014.

1. Nguyen, L. P. & Bradfield, C. A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102–116 (2008).
2. Schrenk, D. Impact of dioxin-type induction of drug-metabolizing enzymes on the metabolism of endo- and xenobiotics. *Biochem. Pharmacol.* **55**, 1155–1162 (1998).
3. Guengerich, F. P. Cytochrome p450 and chemical toxicology. *Chem. Res. Toxicol.* **21**, 70–83 (2008).
4. Casado, F. L., Singh, K. P. & Gasiewicz, T. A. The aryl hydrocarbon receptor: regulation of hematopoiesis and involvement in the progression of blood diseases. *Blood Cells Mol. Dis.* **44**, 199–206 (2010).
5. Marlowe, J. L. & Puga, A. Aryl hydrocarbon receptor, cell cycle regulation, toxicity, and tumorigenesis. *J. Cell. Biochem.* **96**, 1174–1184 (2005).
6. Opitz, C. A. et al. An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478**, 197–203 (2011).
7. Li, Y. et al. Exogenous stimuli maintain intraepithelial lymphocytes via aryl hydrocarbon receptor activation. *Cell* **147**, 629–640 (2011).
8. Lai, Z. W., Pineau, T. & Esser, C. Identification of dioxin-responsive elements (DREs) in the 5' regions of putative dioxin-inducible genes. *Chem. Biol. Interact.* **100**, 97–112 (1996).
9. Jensen, B. A., Leeman, R. J., Schlezinger, J. J. & Sherr, D. H. Aryl hydrocarbon receptor (AhR) agonists suppress interleukin-6 expression by bone marrow stromal cells: an immunotoxicology study. *Environ. Health* **2**, 16 (2003).
10. Kimura, A. et al. Aryl hydrocarbon receptor in combination with Stat1 regulates LPS-induced inflammatory responses. *J. Exp. Med.* **206**, 2027–2035 (2009).

11. Quintana, F. J. *et al.* Control of T_{reg} and TH17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).
12. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links Th17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
13. Apetoh, L. *et al.* The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27. *Nature Immunol.* **11**, 854–861 (2010).
14. Gandhi, R. *et al.* Activation of the aryl hydrocarbon receptor induces human type 1 regulatory T cell-like and Foxp3⁺ regulatory T cells. *Nature Immunol.* **11**, 846–853 (2010).
15. Kiss, E. A. *et al.* Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science* **334**, 1561–1565 (2011).
16. Qiu, J. *et al.* The aryl hydrocarbon receptor regulates gut immunity through modulation of innate lymphoid cells. *Immunity* **36**, 92–104 (2012).
17. Motto, I. *et al.* New aryl hydrocarbon receptor homology model targeted to improve docking reliability. *J. Chem. Inf. Model.* **51**, 2868–2881 (2011).
18. Rada, B. & Leto, T. L. Pyocyanin effects on respiratory epithelium: relevance in *Pseudomonas aeruginosa* airway infections. *Trends Microbiol.* **21**, 73–81 (2013).
19. Kerr, J. R. Phenazine pigments: antibiotics and virulence factors. *Infect Dis. Rev.* **29**, 184–194 (2000).
20. Gardner, P. R. Superoxide production by the mycobacterial and pseudomonad quinoid pigments phthiocol and pyocyanine in human lung cells. *Arch. Biochem. Biophys.* **333**, 267–274 (1996).
21. Lau, G. W., Hassett, D. J., Ran, H. & Kong, F. The role of pyocyanin in *Pseudomonas aeruginosa* infection. *Trends Mol. Med.* **10**, 599–606 (2004).
22. Wilson, R. *et al.* Pyocyanin and 1-hydroxyphenazine produced by *Pseudomonas aeruginosa* inhibit the beating of human respiratory cilia *in vitro*. *J. Clin. Invest.* **79**, 221–229 (1987).
23. Denning, G. M. *et al.* Phenazine-1-carboxylic acid, a secondary metabolite of *Pseudomonas aeruginosa*, alters expression of immunomodulatory proteins by human airway epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **285**, L584–L592 (2003).
24. Caldwell, C. C. *et al.* *Pseudomonas aeruginosa* exotoxin pyocyanin causes cystic fibrosis airway pathogenesis. *Am. J. Pathol.* **175**, 2473–2488 (2009).
25. Morrissey, J. P., Cullinan, M., Abbas, A., Mark, G. L. & O'Gara, F. Biosynthesis and regulation of anti-fungal metabolites by *Pseudomonads*, in *Pseudomonas* Vol III, (ed. by Ramos, J. L.) 637–670 (Springer Verlag, 2004).
26. Hunter, R. C. *et al.* Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. *Am. J. Respir. Cell Mol. Biol.* **47**, 738–745 (2012).
27. Wilson, R. *et al.* Measurement of *Pseudomonas aeruginosa* phenazine pigments in sputum and assessment of their contribution to sputum sol toxicity for respiratory epithelium. *Infect. Immun.* **56**, 2515–2517 (1988).
28. Wincent, E. *et al.* Inhibition of cytochrome P4501-dependent clearance of the endogenous agonist FICZ as a mechanism for activation of the aryl hydrocarbon receptor. *Proc. Natl Acad. Sci. USA* **109**, 4479–4484 (2012).
29. Dietrich, L. E. *et al.* The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*. *Mol. Microbiol.* **61**, 1308–1321 (2006).
30. Reszka, K. J. *et al.* Inactivation of the potent *Pseudomonas aeruginosa* cytotoxin pyocyanin by airway peroxidases and nitrite. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **302**, L1044–L1056 (2012).
31. Muller, M. Glutathione modulates the toxicity of, but is not a biologically relevant reductant for, the *Pseudomonas aeruginosa* redox toxin pyocyanin. *Free Radic. Biol. Med.* **50**, 971–977 (2011).
32. O'Malley, Y. Q. *et al.* *Pseudomonas aeruginosa* pyocyanin directly oxidizes glutathione and decreases its levels in airway epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **287**, L94–L103 (2004).
33. Nebert, D. W., Puga, A. & Vasilou, V. Role of the Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal transduction. *Ann. NY Acad. Sci.* **685**, 624–640 (1993).
34. Lavoie, E. G., Wangdi, T. & Kazmierczak, B. I. Innate immune responses to *Pseudomonas aeruginosa* infection. *Microbes Infect.* **13**, 1133–1145 (2011).
35. Goldberg, J. B. & Pier, G. B. The role of the CFTR in susceptibility to *Pseudomonas aeruginosa* infections in cystic fibrosis. *Trends Microbiol.* **8**, 514–520 (2000).
36. Schuster, M. G. & Norris, A. H. Community-acquired *Pseudomonas aeruginosa* pneumonia in patients with HIV infection. *AIDS* **8**, 1437–1441 (1994).
37. Kobayashi, Y. The role of chemokines in neutrophil biology. *Front. Biosci.* **13**, 2400–2407 (2008).
38. Cheon, H. *et al.* Signaling pathway for 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced TNF- α production in differentiated THP-1 human macrophages. *Exp. Mol. Med.* **39**, 524–534 (2007).
39. Vogel, C. F. & Matsumura, F. A new cross-talk between the aryl hydrocarbon receptor and RelB, a member of the NF- κ B family. *Biochem. Pharmacol.* **77**, 734–745 (2009).
40. Eisele, N. A. & Anderson, D. M. Host defense and the airway epithelium: frontline responses that protect against bacterial invasion and pneumonia. *J. Pathogens* <http://dx.doi.org/10.4061/2011/249802> (2011).
41. Lee, Y. C. *et al.* 2,3,7,8-Tetrachlorodibenzo-p-dioxin-induced MUC5AC expression: aryl hydrocarbon receptor-independent/EGFR/ERK/p38-dependent SP1-based transcription. *Am. J. Respir. Cell Mol. Biol.* **45**, 270–276 (2011).
42. Dorhoi, A., Reece, S. T. & Kaufmann, S. H. For better or for worse: the immune response against *Mycobacterium tuberculosis* balances pathology and protection. *Immunol. Rev.* **240**, 235–251 (2011).
43. Weiner, J., III *et al.* Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS ONE* **7**, e40221 (2012).
44. Mavrodi, D. V., Blankenfeldt, W. & Thomashow, L. S. Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. *Annu. Rev. Phytopathol.* **44**, 417–445 (2006).
45. Pierson, L. S., III & Pierson, E. A. Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl. Microbiol. Biotechnol.* **86**, 1659–1670 (2010).
46. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
47. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A., Jr. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are highly grateful to J. Welch for his contribution in the early stage of research, and for support in AhR ligand screening. A. Zychlinsky and L. E. Dietrich for discussions, H.-G. Hoymann for support in lung function measurements and interpretation of results, U. Klemm for mouse breeding and M. L. Grossman for excellent support in preparing the manuscript. *Ahr*^{−/−} mice were provided by B. Stockinger and shRNA constructs by D. Krastev and F. Buchholz. The PA14 WT2 and PA14 Δ phz1/2 were a gift from D. K. Newman and L. E. Dietrich. M. Kolbe acknowledges grant support from the European Union's Seventh Framework Programmes (EU-FP7/2007–2013). S.H.E.K. acknowledges grant support from the European Union's Seventh Framework Programmes (EU-FP7) NEWTBVAC (Health-F3-2009-241745) and PHAGOSYS (Health-F4-2008-223451).

Author Contributions P.M.-A., E.H., K.F., A. Dorhoi and S.H.E.K. conceived and designed the study and wrote the manuscript. P.M.-A., E.H., K.F. and A. Dorhoi designed and performed experiments and data analysis. P.C. and M.D. performed *Mycobacterium* lipid fractionation. A.M. and B.T. designed and performed *Pseudomonas* spirometry studies. U.G.-B., M. Klemm, A.-B.K. and S.B. provided technical help for *in vitro* and *in vivo* experiments. R.H. performed and analysed HPLC experiments. A.K., G.K. and H.O. performed virtual docking studies. S.F. discussed experiments. T.S., V.B. and C.G. performed and discussed experiments. H.-J.M. performed and analysed microarray experiments. M. Kolbe, N.B., J.F., A. Diehl and H.O. performed binding studies. All authors commented on the paper.

Author Information Data are deposited in the GEO database under accession number GSE48133. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H.E.K. (kaufmann@mpiib-berlin.mpg.de).

Diversity and dynamics of the *Drosophila* transcriptome

James B. Brown^{1,2*}, Nathan Boley^{1*}, Robert Eisman^{3*}, Gemma E. May^{4*}, Marcus H. Stoiber^{1*}, Michael O. Duff⁴, Ben W. Booth², Jiayu Wen⁵, Soo Park², Ana Maria Suzuki^{6,7}, Kenneth H. Wan², Charles Yu², Dayu Zhang⁸, Joseph W. Carlson², Lucy Cherbass³, Brian D. Eads³, David Miller³, Keithanne Mockaitis³, Johnny Roberts⁸, Carrie A. Davis⁹, Erwin Frise², Ann S. Hammonds², Sara Olson⁴, Sol Shenker⁵, David Sturgill¹⁰, Anastasia A. Samsonova^{11,12}, Richard Weiszmann², Garret Robinson¹, Juan Hernandez¹, Justen Andrews³, Peter J. Bickel¹, Piero Carninci^{6,7}, Peter Cherbass^{3,8}, Thomas R. Gingeras⁹, Roger A. Hoskins², Thomas C. Kaufman³, Eric C. Lai⁵, Brian Oliver¹⁰, Norbert Perrimon^{11,12}, Brenton R. Graveley⁴ & Susan E. Celniker²

Animal transcriptomes are dynamic, with each cell type, tissue and organ system expressing an ensemble of transcript isoforms that give rise to substantial diversity. Here we have identified new genes, transcripts and proteins using poly(A)⁺ RNA sequencing from *Drosophila melanogaster* in cultured cell lines, dissected organ systems and under environmental perturbations. We found that a small set of mostly neural-specific genes has the potential to encode thousands of transcripts each through extensive alternative promoter usage and RNA splicing. The magnitudes of splicing changes are larger between tissues than between developmental stages, and most sex-specific splicing is gonad-specific. Gonads express hundreds of previously unknown coding and long non-coding RNAs (lncRNAs), some of which are antisense to protein-coding genes and produce short regulatory RNAs. Furthermore, previously identified pervasive intergenic transcription occurs primarily within newly identified introns. The fly transcriptome is substantially more complex than previously recognized, with this complexity arising from combinatorial usage of promoters, splice sites and polyadenylation sites.

Next-generation RNA sequencing (RNA-seq) has permitted the mapping of transcribed regions of the genomes of a variety of organisms^{1,2}. These studies demonstrated that large fractions of metazoan genomes are transcribed, and they also catalogued individual elements of transcriptomes, including transcription start sites³, polyadenylation sites^{4,5}, exons and introns⁶. However, the complexity of the transcriptome arises from the combinatorial incorporation of these elements into mature transcript isoforms. Studies that inferred transcript isoforms from short-read sequence data focused on a small subset of isoforms, filtered using stringent criteria^{7,8}. Studies using complementary DNA (cDNA) or expressed sequence tag (EST) data to infer transcript isoforms have not had sufficient sampling depth to explore the diversity of RNA products at most genomic loci⁹. Although the human genome has been the focus of intensive manual annotation¹⁰, analysis of strand-specific RNA-seq data from human cell lines reveals over 100,000 splice junctions not incorporated into transcript models¹¹. Thus, a large gap exists between genome annotations and the emerging transcriptomes observed in next-generation sequence data. In *Drosophila*, we previously described a non-strand-specific poly(A)⁺ RNA-seq analysis of a developmental time course through the life cycle⁶ and cap analysis of gene expression (CAGE) analysis of the embryo¹², which discovered thousands of unannotated exons, introns and promoters, and expanded coverage of the genome by identified transcribed regions, but not all elements were incorporated into full-length transcript models. Here we describe an expansive poly(A)⁺ transcript set modelled by integrative analysis of transcription

start sites (CAGE and 5' rapid amplification of cDNA ends (RACE)), splice sites and exons (RNA-seq), and polyadenylation sites (3' expressed sequence tags (ESTs), cDNAs and RNA-seq). We analysed poly(A)⁺ RNA data from a diverse set of developmental stages⁶, dissected organ systems and environmental perturbations; most of this data is new and strand-specific. Our data provide higher spatiotemporal resolution and allow for deeper exploration of the *Drosophila* transcriptome than was previously possible. Our analysis reveals a transcriptome of high complexity that is expressed in discrete, tissue- and condition-specific messenger RNA and lncRNA transcript isoforms that span most of the genome and provides valuable insights into metazoan biology.

A dense landscape of discrete poly(A)⁺ transcripts

To broadly sample the transcriptome, we performed strand-specific, paired-end sequencing of poly(A)⁺ RNA in biological duplicate from 29 dissected tissue samples including the nervous, digestive, reproductive, endocrine, epidermal and muscle organ systems of larvae, pupae and adults. To detect RNAs not observed under standard conditions, we sequenced poly(A)⁺ RNA in biological duplicate from 21 whole-animal samples treated with environmental perturbations. Adults were challenged with heat-shock, cold-shock, exposure to heavy metals (cadmium, copper and zinc), the drug caffeine or the herbicide paraquat. To determine whether exposing larvae resulted in RNA expression from previously unidentified genes, we treated them with heavy metals, caffeine, ethanol or rotenone. Finally, we sequenced poly(A)⁺ RNA from 21

¹Department of Statistics, University of California Berkeley, Berkeley, California 94720, USA. ²Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Department of Biology, Indiana University, 1001 East 3rd Street, Bloomington, Indiana 47405, USA. ⁴Department of Genetics and Developmental Biology, Institute for Systems Genomics, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, Connecticut 06030, USA. ⁵Sloan-Kettering Institute, 1017C Rockefeller Research Labs, 1275 York Avenue, Box 252, New York, New York 10065, USA. ⁶RIKEN Omics Science Center, Yokohama, Kanagawa 230-0045, Japan. ⁷RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, 230-0045, Japan. ⁸Center for Genomics and Bioinformatics, Indiana University, 1001 East 3rd Street, Bloomington, Indiana 47405, USA. ⁹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ¹⁰Section of Developmental Genomics, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹¹Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ¹²Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

previously described¹³ and three ovary-derived cell lines (Supplementary Methods). In total, we produced 12.4 billion strand-specific read pairs and over a terabase of sequence data, providing 44,000-fold coverage of the poly(A)⁺ transcriptome.

Reads were aligned to the *Drosophila* genome as described⁶, and full-length transcript models were assembled using our custom pipeline termed GRIT¹⁴, which uses RNA-seq, poly(A)⁺seq, CAGE, RACE¹², ESTs¹⁵ and full-length cDNAs¹⁶ to generate gene and transcript models (Supplementary Methods). We integrated these models with our own and community manual curation data sets to obtain an annotation (Supplementary Information, section 12) consisting of 304,788 transcripts and 17,564 genes (Fig. 1a and Supplementary Fig. 1), of which 14,692 are protein-coding (Supplementary Data 1 and updates available at <http://fruitfly.org>). Ninety per cent of genes produce at most 10 transcript and five protein isoforms, whereas 1% of genes have highly complex patterns of alternative splicing, promoter usage and polyadenylation, and may each be processed into hundreds of transcripts (Fig. 1a, b). Our gene models span 72% of the euchromatin, an increase from 65% in FlyBase 5.12 (FB5.12), the reference annotation at the beginning of the modENCODE project (Supplementary Table 1 compares annotations in 2008–13). There were 64 euchromatic gene-free regions longer than 50 kb in FB5.12, and 25 remaining in FB5.45. Our annotation includes new gene models in each of these regions. Newly identified genes (1,468 total) are expressed in spatially and temporally restricted patterns (Supplementary Fig. 2), and 536 reside in previously uncharacterized gene-free regions. Others map to well-characterized regions, including the *ovo* locus, where we discovered a new ovary-specific, poly(A)⁺ transcript (*Mgn94020*, Supplementary Data 1 and 2), extending

from the second promoter of *ovo* on the opposite strand and spanning 107 kb (Fig. 1c). Exons of 36 new genes overlap molecularly defined mutations with associated phenotypes (genome structure correction (GSC) *P* value ~0.0002), indicating potential functions (Supplementary Table 2). For example, the lethal P-element insertions *l(3)L3051* and *l(3)L4111* (ref. 17) map to promoters of *Mgn095159* and *Mgn95009*, respectively, indicating these may be essential genes. Nearly 60% of the intergenic transcription we previously reported⁶ is now incorporated into gene models.

Transcript diversity

Over half of spliced genes (7,412; 56%) encode two or more transcript isoforms with alternative first exons. Most of such genes produce alternative first exons through coordinated alternative splicing and promoter usage (59%, 4,389 genes, hypergeometric *P* value < 1×10^{-16}); however, a substantial number of genes use one, but not both mechanisms (Fig. 2a). Only 1,058 spliced genes have alternative first exons that alter protein-encoding capacity and increase the complexity of the predicted proteome. Some genes, such as *G protein β -subunit 13F* (*G β 13F*, Fig. 2b and Supplementary Fig. 3) have exceptionally complex 5' UTRs, but encode a single protein.

We measured splicing efficiency using the 'per cent spliced in' (Ψ) index—the fraction of isoforms that contain the particular exon⁶. Introns flanked by coding sequence are retained at an average $\Psi = 0.7$, whereas introns flanked by non-coding sequence are retained > fivefold more often, with an average $\Psi = 3.8$ ($P < 1 \times 10^{-16}$ subsampling/two-sample *t*-test), and is most frequent in 5' UTRs (mean $\Psi = 5.1$, Fig. 2c).

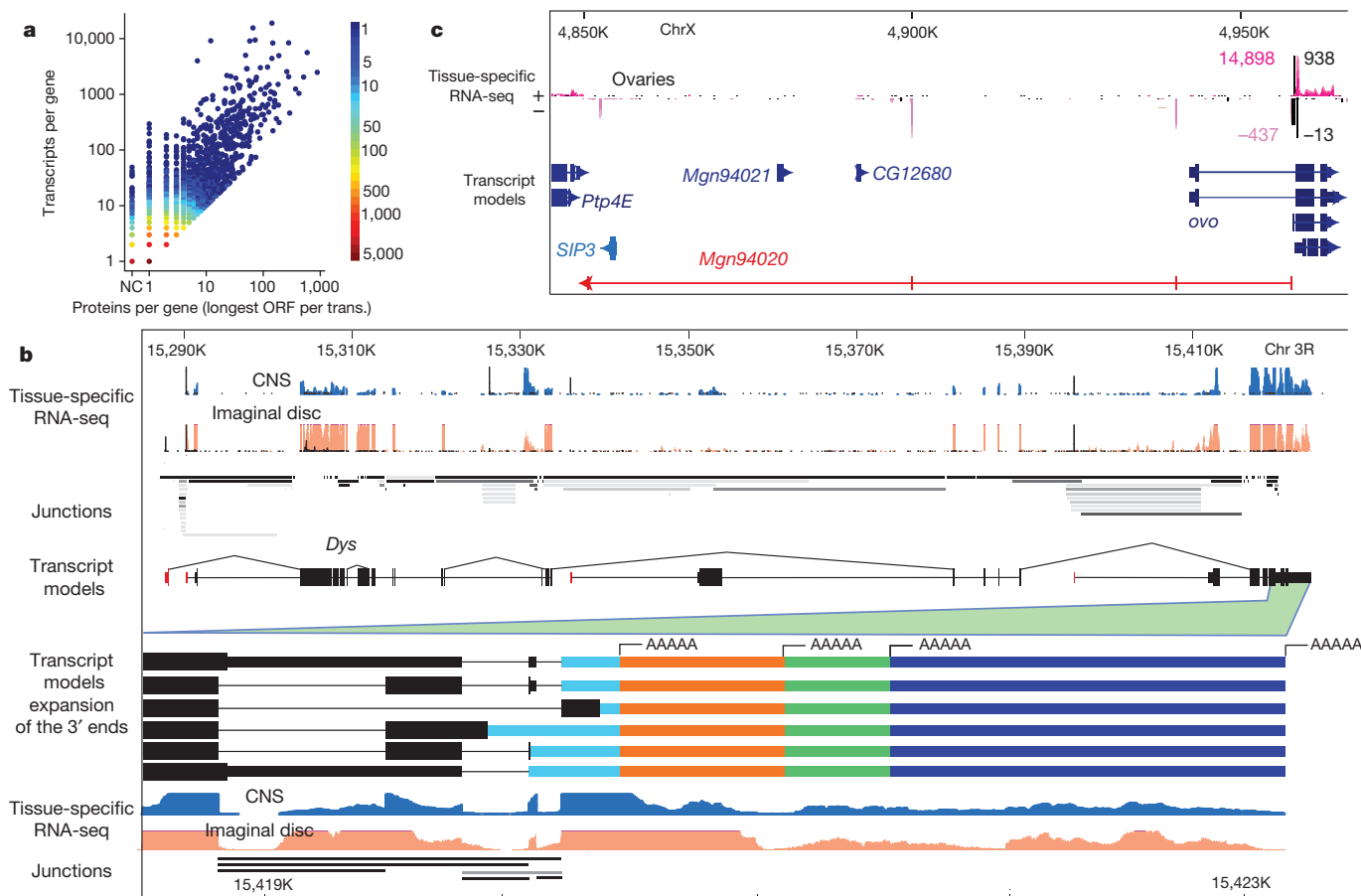


Figure 1 | Overview of the annotation of the *Drosophila melanogaster* transcriptome. **a**, Scatterplot showing the per gene correlation between number of proteins and number of transcripts. The genes *Dscam* and *para* are omitted as extreme outliers both encoding >10,000 unique proteins. **b**, *Dystrophin* (*Dys*) produces 72 transcripts and encodes 32 proteins.

Highlighted is alternative splicing and polyadenylation at the 3' end. CAGE (black), RNA-seq (tan, blue), splice junctions (shaded grey as a function of usage). **c**, An internal promoter of *ovo* is bidirectional in ovaries and produces a lncRNA (430 bp, red) bridging two gene deserts. CAGE (black), RNA-seq (pink), counts are read-depth (minus-strand given as negative).

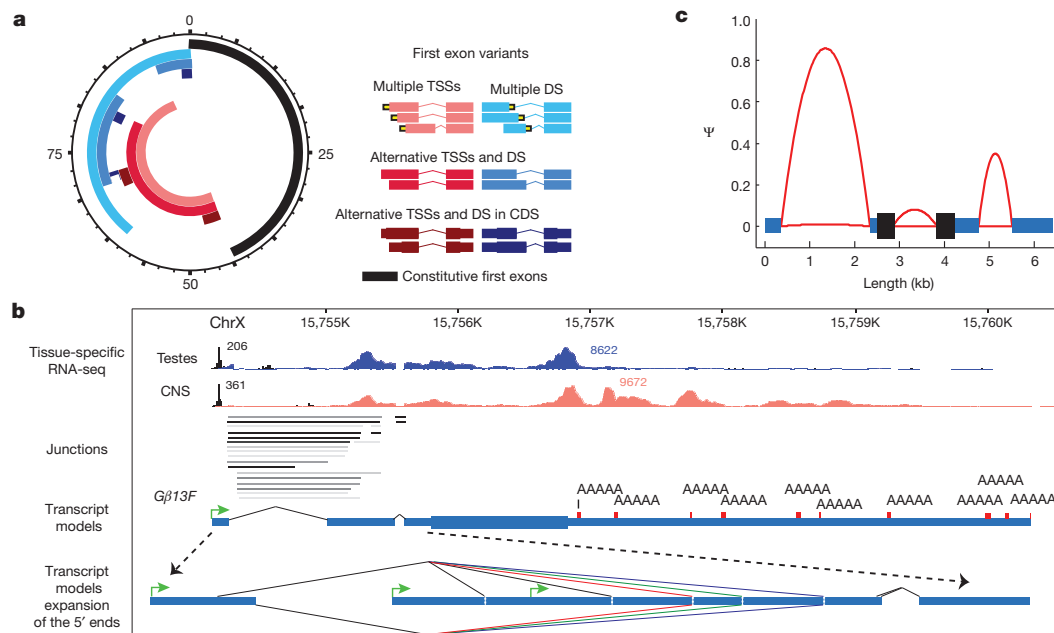


Figure 2 | Splicing complexity across the gene body. **a**, Alternative first exons occur in two main configurations: multiple transcription start sites (TSS, pink) and multiple donor sites (DS, light blue). A subset of the genes in the multiple TSS category produce transcripts with different TSSs and shared donor sites (red), and a subset of the genes in the DS category produce transcripts with a shared TSS and different donor sites (blue). Some genes in the multiple TSS category directly affect the encoded protein (maroon), and similarly for DS (dark blue). The overlap of configurations is radially proportional (units indicate percentage of all spliced genes). **b**, Poly(A)⁺ testes (blue) and central nervous system (CNS) (orange) stranded RNA-seq of *Gβ13F* showing complex processing and splicing of the 5' UTR. An expansion of the 5' UTR showing some of the complexity. Transcription of the gene initiates from one of three different promoters (green arrows) terminates at one of ten possible poly(A)⁺

addition sites (from adult head poly(A)⁺ seq, red) and generates 235 transcripts. The first exon has two alternative splice acceptors that splice to one of eleven different donor sites. Only five donor sites are shown owing to the proximity of splice sites. Four splice donors are represented by the single red line differing by 12, 5 and 19 bp, respectively. Three splice donors are represented by the single green line differing by 12 and 11 bp. Two splice donors are represented by the single purple line differing by 7 bp. These splice variants are combined with four proximal internal splices (Supplementary Fig. 3a) to generate the full complement of transcripts. **c**, Intron retention rates (Ψ) across the gene body. The genome-wide mean lengths of exons and introns are connected by red parabolic arcs, which illustrate the upper and lower quartiles of intron retention (across all samples) for introns retained at or above 20 Ψ in at least one sample.

Despite the depth of our RNA-seq, these data show that 42% of genes encode only a single transcript isoform, and 55% encode a single protein isoform (Supplementary Methods). In mammals, it has been estimated that 95% of genes produce multiple transcript isoforms^{18,19}, (estimates for protein-coding capacity have not been reported).

The majority of transcriptome complexity is attributable to forty-seven genes that have the capacity to encode >1,000 transcript isoforms each (Supplementary Table 3), and account for 50% of all transcripts (Fig. 3a). Furthermore, 27% of transcripts encoded by these genes were detected exclusively in samples enriched for neuronal tissue, and another 56% only in the embryo (83% total). To determine their tissue specificities we conducted embryonic *in situ* expression assays (Fig. 3b) and found that 18 of 35 are detected only in neural tissue (51% compared with 10% genome-wide, hypergeometric P value $< 1 \times 10^{-16}$, Supplementary Table 4). Of these genes, 48% have 3' UTR extensions in embryonic neural tissue²⁰ (5% genome-wide, $P < 1 \times 10^{-16}$). Furthermore,

44% are targets of RNA editing (4% genome-wide⁶, $P < 1 \times 10^{-16}$, with 18 of 21 validated²¹), and 21% have 3' UTR extensions and RNA editing sites (10 of 65 genome-wide, $P < 1 \times 10^{-100}$). The capacity to encode thousands of transcripts is largely specific to the nervous system and coincides with other classes of rare, neural-specific RNA processing.

Tissue- and sex-specific splicing

To examine the dynamics of splicing, we calculated switch scores or $\Delta\Psi$, for each splicing event by comparing the maximal and minimal Ψ values across all samples, and in subsets including just the developmental and tissue samples. In contrast to the median Ψ values, the distribution of $\Delta\Psi$ values is strikingly different between the developmental and tissue samples. Among the developmental samples, 38% of events have a $\Delta\Psi \geq 50\%$, whereas between the tissue samples 63% of events have a $\Delta\Psi \geq 50\%$. This difference is even more pronounced

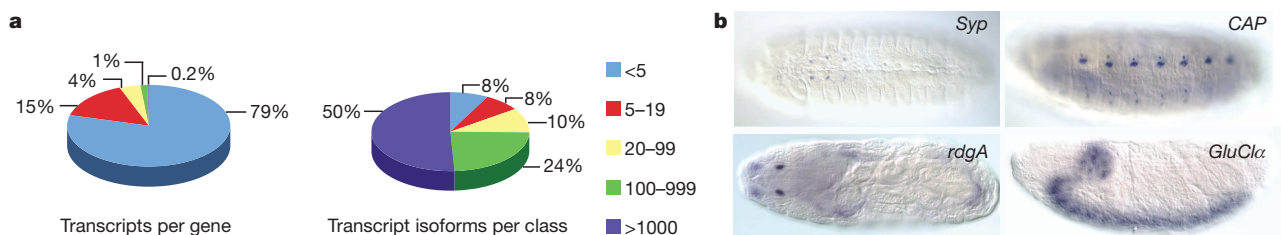


Figure 3 | Complex splicing patterns are mainly limited to neural tissues. **a**, A small minority of genes (47, 0.2%) encode most transcripts. **b**, *In situ* RNA staining of constitutive exons of four genes with highly complex splicing patterns in the embryo. *Syncrip* (*Syp*), *CAP*, *retinal degeneration A* (*rdgA*) and

GluClα show specific late embryonic neural expression in the ventral midline neurons; dorsal/lateral and ventral sensory complexes; Bolwig's organ or larval eye; and central nervous system, respectively.

at higher $\Delta\Psi$ thresholds—only 6% of events have a $\Delta\Psi \geq 80\%$ between the developmental samples, whereas 31% of events have a $\Delta\Psi \geq 80\%$ between the tissue samples. Thus, most splicing events are highly tissue-specific. Of the 17,447 alternative splicing events analysed (Supplementary Information, section 19), we find that 56.6% changed significantly ($\Delta\Psi > 20\%$, Bayes factor > 20). Clustering revealed groups of splicing events that are co-ordinately regulated in a tissue-specific manner. For example, 1,147 splicing events are specifically included in heads and excluded in testes or ovaries, whereas 797 splicing events are excluded in heads but included in testes or ovaries (Fig. 4a).

We identified hundreds of sex-specific splicing events from adult male and female RNA-seq data⁶. To further explore sex-specific splicing, we compared the splicing patterns in male and female heads enriched for brain tissues. There were striking differences in gene expression levels, however, only seven splicing events were consistently differentially spliced at each time point after eclosion (average $\Delta\Psi > 20\%$), and these largely corresponded to genes in the known sex-determination pathway (Supplementary Information, section 19A). We find few examples of head sex-specific splicing. This is in contrast to previous studies, which have come to conflicting conclusions and used either microarrays analysing only a subset of splicing events or single read 36-bp RNA-seq^{22,23} with an order of magnitude fewer reads²⁴.

We identified 575 alternative splicing events that are differentially spliced in whole male and female animals ($\Delta\Psi > 20\%$) and analysed the tissue-specific splicing patterns of each event (Fig. 4b). We found that 186 of the 321 male-biased splicing events were most strongly included in testes or accessory glands, and 157 of 254 female-biased exons were ovary-enriched. Consistent with the extensive transcriptional differences observed in testes compared to other tissues, the genes containing male-specific exons are enriched in functions related to transcription. In contrast, the female-specific exon containing genes are enriched in functions involved in signalling and splicing (<http://reactome.org>)²⁵, Supplementary Table 6). Together, these results indicate that the majority of sex-specific splicing is due to tissue-specific splicing in tissues present only in males or females.

Long non-coding RNAs

A growing set of candidate long non-coding RNAs (lncRNAs) have been identified in *Drosophila*^{6,26,27}. In FB5.45 there were 392 annotated lncRNAs, and it has been suggested that as many as 1,119 lncRNAs may be transcribed in the fly²⁸. However, this number was based on transcribed regions, not transcript models, and used non-stranded RNA-seq data²⁸. We find 3,880 genes produce transcripts with ORFs encoding fewer than 100 amino acids. Of these, 795 encode conserved proteins (Methods) longer than 20 amino acids. For example, a single exon gene on the opposite strand and in the last intron of the early developmental growth factor *spätzle* encodes a 42-amino-acid protein that is highly conserved across all sequenced *Drosophila* species. We identified 1,875 candidate lncRNA genes producing 3,085 transcripts, 2,990 of which have no overlap with protein-coding genes on the same strand (Supplementary Data 2). Some of these putative lncRNAs may encode short polypeptides, for example, the gene *tarsal-less* encodes three 11-amino-acid ORFs with important developmental functions²⁹. We determined protein conservation scores for each ORF between 20 and 100 amino acids (Supplementary Table 6). Of the 1,119 predicted lncRNAs²⁸, we provide full-length transcript models for 246 transcribed loci; the remainder were expressed at levels beneath thresholds used in this study. This is not surprising, the expression patterns of lncRNAs are more restricted than those of protein-coding genes: the average lncRNA is expressed (bases per kilobase per million mapped bases⁶ (BPKM) > 1) in 1.5 developmental and 3.2 tissue samples, compared to 6.6 and 17 for protein-coding genes, respectively. Many lncRNAs (563 or 30%) have peak expression in testes, and 125 are detectable only in testes. Similarly restricted expression patterns have been reported for lncRNAs in humans and other mammals^{30,31}.

Interestingly, all newly annotated genes overlapping molecularly defined mutations with phenotypes are lncRNAs (Supplementary Table 2). For instance, the mutation D114.3 is a regulatory allele of *spineless* (*ss*) that maps 4 kb upstream of *ss*³² and within the promoter of *Mgn4221*. Similarly, *Mgn00541* corresponds to a described, but unannotated 2.0 kb transcript overlapping the regulatory mutant allele *ci*⁵⁷ of *cubitus*

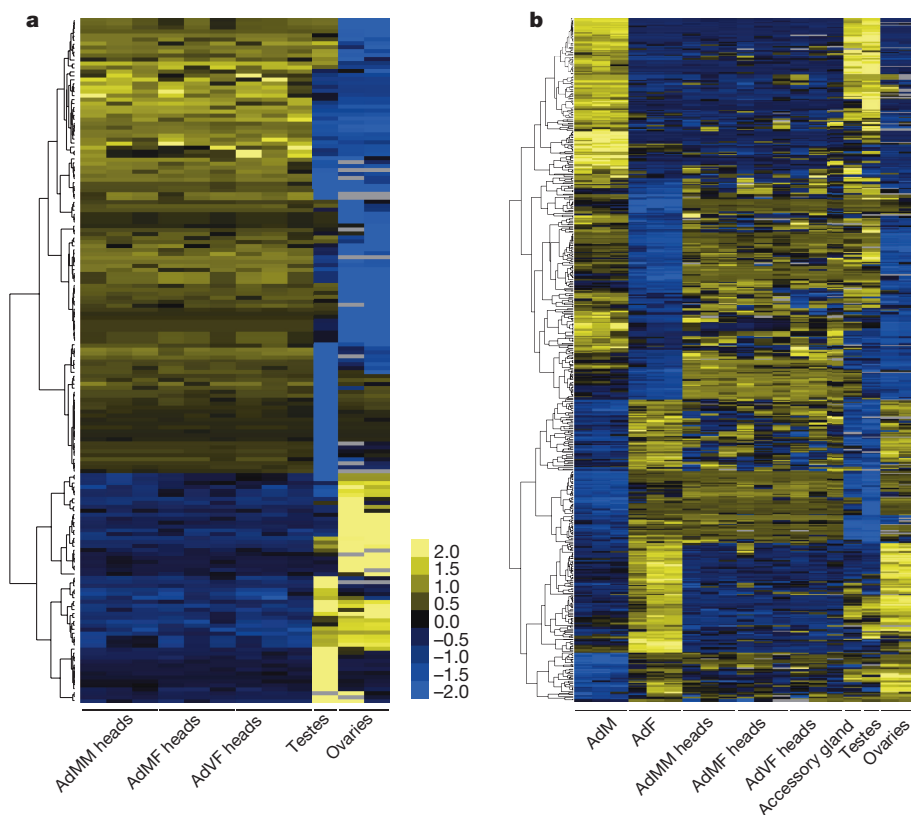


Figure 4 | Sex-specific splicing is mainly tissue-specific splicing. **a**, Clusters of tissue-specific splicing events. The scale bar indicates z-scores of Ψ . Adult mated male (AdMM), adult mated female (AdMF) and adult virgin female (AdVF) heads are from 1-, 4- and 20-day-old animals, respectively. Testes are from 4-day-old adult males, and ovaries are from mated and virgin 4-day-old adult females. **b**, Sex-specific splicing events in whole animals are primarily testes- or ovary-specific splicing events. Adult male (AdM) and adult female (AdF) animals are 1, 5 and 30 days old. Accessory glands were dissected from 4-day-old adult males. The RNA-seq columns from heads, testes and ovaries are as described in a.

*interruptus*³³. It remains to be determined whether these mutations are a result of the loss-of-function of newly annotated transcripts or *cis*-acting regulatory elements (for example, enhancers) or both.

Antisense transcription

Drosophila antisense transcription has been reported³⁴, but the catalogue of antisense transcription has been largely limited to overlapping mRNAs transcribed on opposite strands. We identify non-coding antisense transcript models for 402 lncRNA loci that are antisense to mRNA transcripts of 422 protein-coding genes (for example, *prd*, Fig. 5a), and 36 lncRNAs form 'sense-antisense gene-chains' overlapping more than one protein-coding locus, as observed in mammals^{30,35}. In *Drosophila*, 21% of lncRNAs are antisense to mRNAs, whereas in human 15% of annotated lncRNAs are antisense to mRNAs (GENCODE v.10). We assembled antisense transcript models for 5,057 genes (29%, compared to previous estimates of 15%³⁴). For 67% of these loci, antisense expression is observable in at least one cell line, indicating that sense/antisense transcripts may be present in the same cells. lncRNA-mediated antisense accounts for a small minority of antisense transcription: 94% of antisense loci correspond to overlapping protein-coding mRNAs transcribed on opposite strands, and of these, 323 loci (667 genes) share overlapping CDSs. The majority of antisense is due to overlapping UTRs: 1,389 genes have overlapping 5' UTRs (divergent transcription), 3,430 have overlapping 3' UTRs (convergent transcription), and 540 have both, meaning that, as with many lncRNAs, they form gene-chains across contiguously transcribed regions. A subset of antisense gene-pairs overlap almost completely (>90%), which we term reciprocal transcription. There are 13 such loci (Supplementary Fig. 5) and seven are male-specific (none are female-specific).

The mRNA/lncRNA sense-antisense pairs tend to be more positively correlated in their expression than mRNA/mRNA pairs, (mean $r = 0.16$ compared with 0.13, Kolmogorov–Smirnov (KS) two-sample one-sided test $P < 10^{-9}$), and although this mean effect is subtle, the trend is clearly visible in the quantiles (95th percentile lncRNA/mRNA 0.729 versus mRNA/mRNA 0.634, Supplementary Fig. 6a). This effect is stronger when the analysis is restricted to cell line samples (Supplementary Fig. 6b).

Even in homogenous cell cultures, evidence for sense-antisense transcription does not guarantee that both transcripts exist within individual cells: transcription could originate from exclusive events occurring in different cells. *Cis*-natural antisense transcripts (*cis*-NATs) are a substantial source of endogenous siRNAs³⁶, and their existence directly reflects the existence of precursor dsRNA. *Cis*-NAT-siRNA production typically involves convergent transcription units that overlap on their 3' ends, but other documented loci generate siRNAs across internal exons, introns or 5' UTRs^{37–39}. Analysis of head, ovary and testis RNAs showed that 328 unique sense/antisense gene pair regions generated 21-nucleotide RNAs indicative of siRNA production (Supplementary Table 8), and these were significantly enriched (Supplementary Fig. 7a, Supplementary Methods) for pairs showing positively correlated expression between sense and antisense levels across tissues ($P = 2 \times 10^{-5}$), embryo developmental stages ($P = 4 \times 10^{-3}$), conditions ($P = 9 \times 10^{-4}$)

and across all samples ($P = 3 \times 10^{-5}$). The tissue distribution of these *cis*-NAT-siRNAs showed a bias for testis expression (Supplementary Fig. 7b), with fourfold greater number relative to ovaries ($P = 2 \times 10^{-17}$, binomial test) and sevenfold relative to heads ($P = 4 \times 10^{-25}$) and expression levels of siRNAs were substantially higher in testes than other tissues (Supplementary Fig. 7c).

Over 80% of *cis*-NAT-siRNAs were derived from 3'-convergent gene pairs. Abundant siRNAs emanate from an overlap of the *gryzun* and *CG14967* 3' UTRs (Supplementary Fig. 5). The remainder were distributed amongst CDSs, introns and 5' UTRs. We identified abundant testis-enriched siRNA production from a 5'-divergent overlap of *Cyt-c-d* and *CG31808* (Fig. 5b) and from the entire CDS of *dUTPase* and its antisense non-coding transcript *Mgn99994*.

Transcriptional effects of environmental stress

Whole-animal perturbations each exhibited condition-specific effects, for example, the metallothionein genes were induced by heavy metals (Fig. 6a), but not by other treatments (Supplementary Table 9). The genome-wide transcriptional response to cadmium (Cd) exposure involves small changes in expression level in thousands of genes (48 h after exposure), but only a small group of genes change > 20-fold, and this group includes six lncRNAs (the third most strongly induced gene is *CR4138*, Fig. 6a, Supplementary Fig. 8a). Four newly modelled lncRNAs are differentially expressed (1% false discovery rate (FDR)) in at least one treatment, and constitute newly described eco-responsive genes. Furthermore, 57 genes and 5,259 transcripts (of 811 genes) were detected exclusively in these treatment samples. Although no two perturbations revealed identical transcriptional landscapes, we find a homogeneous response to environmental stressors (Fig. 6b, Supplementary Fig. 8b). The direction of regulation for most genes is consistent across all treatments; very few are upregulated in one condition and downregulated in another. Classes of strongly upregulated genes included those annotated with the GO term "Response to Stimulus, GO:0050896" (most enriched, P value $< 1 \times 10^{-16}$, Supplementary Fig. 8c), and those that encode lysozymes (> tenfold), cytochrome P450s, and mitochondrial components *mt:ATPase6*, *mt:CoI*, *mt:CoIII* (> fivefold). Genes encoding egg-shell, yolk and seminal fluid proteins are strongly downregulated in response to every treatment except 'cold2' and 'heat shock' (Supplementary Fig. 8d). For these two stressors, samples were collected 30 min after exposure, corresponding to an 'early response test' showing suppression of germ cell production is not immediate.

Discussion

Most transcriptional complexity in *Drosophila* occurs in tissues of the nervous system, and particularly in the functionally differentiating central and peripheral nervous systems. A subset of ultra-complex genes encodes more than half of detected transcript isoforms and these are dramatically enriched for RNA editing events and 3' UTR extensions, both phenomena largely specific to the nervous system. Our study indicates that the total information output of an animal transcriptome may be heavily weighted by the needs of the developing nervous system.

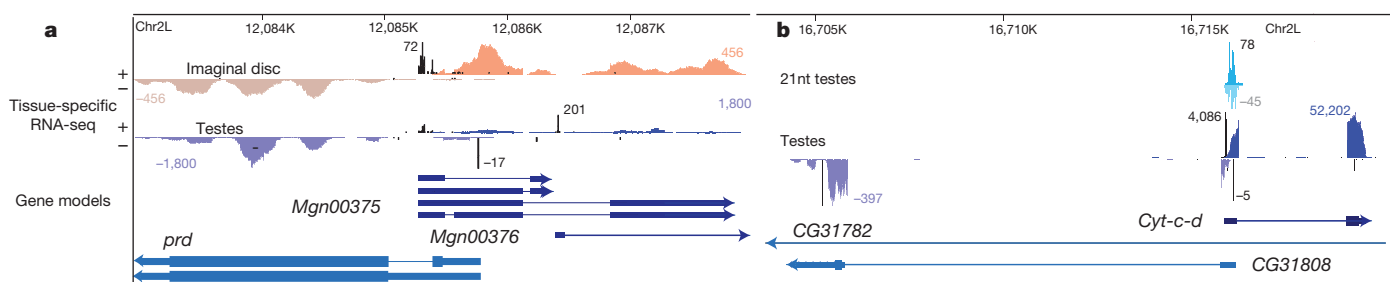


Figure 5 | Examples of antisense transcription. a, 5'/5' overlapping bidirectional antisense transcription at the *prd* locus. Short RNA sequencing does not reveal substantial siRNA (that is, 21-nt-dominant small RNA) signal

in this region (data not shown). b, A 5'/5' antisense region that produces substantial small RNA signal on both strands. nt, nucleotide.

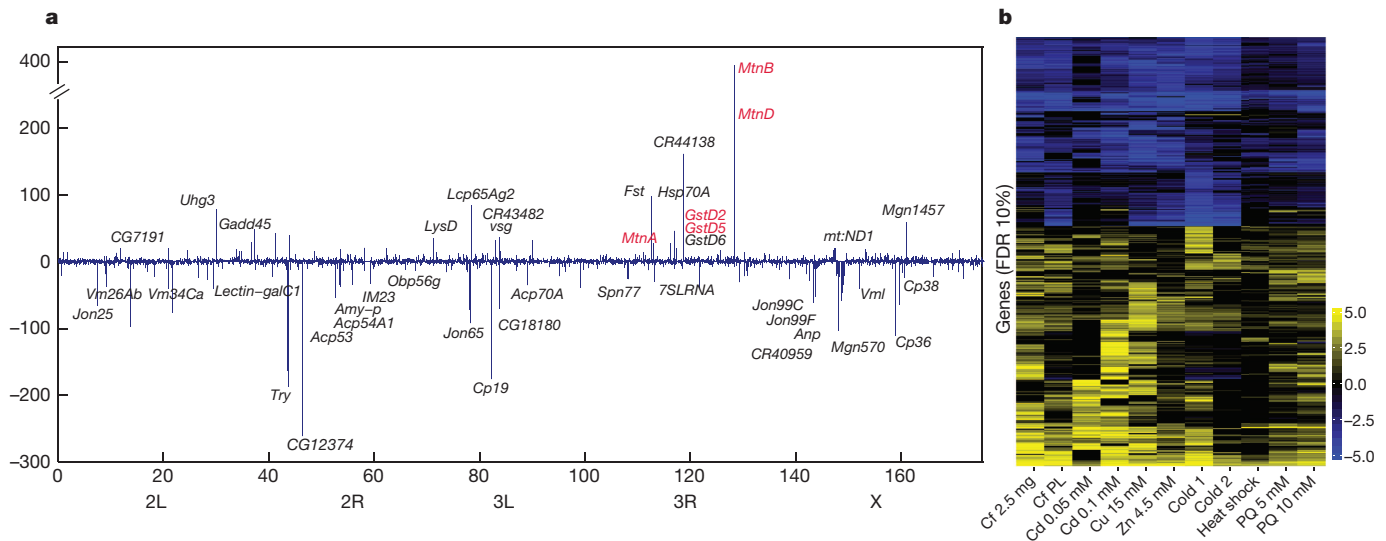


Figure 6 | Effects of environmental perturbations on the *Drosophila* transcriptome. Adults were treated with caffeine (Cf), Cd, Cu, Zn, cold, heat or paraquat (PQ). **a**, A genome-wide map of genes that are up- or downregulated as a function of Cd treatment. Labeled genes are those that showed a 20-fold (<10% FDR) change in response (linear scale). Genes highlighted in red are

The improved depth of sampling and spatiotemporal resolution resulted in the identification of more than 1,200 new genes not discovered in our previous study of *Drosophila* development⁶. A large fraction of the new genes are testes-specific, and many of these are antisense RNAs, as previously described in mammals³⁰. Some new lncRNAs, such as *Mgn94020* (Fig. 1), form sense/antisense gene-chains that bring distant protein-coding genes into transcriptional relationships, another phenomenon previously described only in mammals⁴⁰. Whenever *Mgn94020* is detectably transcribed, the genes on the opposite strand in its introns are not, indicating that its transcription may serve a regulatory function independent of the RNA transcribed. The presence of short RNAs at many regions of antisense transcription indicates that sense and antisense transcripts are present in the same cells at the same times. Many of these *Drosophila* antisense transcripts correspond to 'positionally equivalent'³⁰ antisense transcripts in human. In the two species we found antisense lncRNAs opposite to orthologous protein-coding genes. The apparent positional equivalence of fly and human antisense transcription at genes like *Monocarboxylate transporter 1* (*MCT1*), *even-skipped* (*EVX1*), *CTCF* (*CTCF*), *Adenosine receptor* (*ADORA2A*), and many others^{10,31} across 600 million years of evolution suggests a conserved regulatory mechanism basal to sexual reproduction in metazoans.

Perturbation experiments identified new genes and transcripts, but perhaps more importantly, a general response to stress that is broader than the heat shock pathway. A similar study conducted on marsh fishes in the wake of the Deepwater Horizon incident in the Gulf of Mexico⁴¹ demonstrated that the killifish response to chronic hydrocarbon exposure included induction of lysosome genes, P450 cytochromes and mitochondrial components, and the downregulation of genes encoding eggshell and yolk proteins⁴¹. This overlap of expression responses by gene families across phyla suggests a conserved metazoan stress response involving enhanced metabolism and the suppression of genes involved in reproduction.

We defined an extensive catalogue of putative lncRNAs. However, many genes are known to encode poorly conserved, short polypeptides, including genes specific to the male gonad and accessory gland. Analysis of ribosome profiling initially indicated that a number of mammalian lncRNAs may be translated⁴², but this observation has been difficult to validate by proteomics⁴³, and further analysis has suggested that although lncRNAs have signatures of ribosome occupancy,

those identified in larvae⁵⁰. Some genes are omitted for readability, the complete figure and list of omitted genes are given in Supplementary Fig. 8a. **b**, Heat map showing the fold change of genes with a FDR < 10% (differential expression) in at least one sample (log₂ scale). PL, pre-lethal.

they are not translated⁴⁴. Therefore, while we refer to these RNAs as 'non-coding', additional data are needed to determine if they produce small polypeptides.

The biological consequences of many of the phenomena reported here, including the observation that many genes encoding RNA binding proteins exhibit extraordinary splicing complexity, often within their 5' UTRs, require further study. The splicing factor *pUf68* encodes more than 100 alternatively spliced 5' UTR variants, but encodes a single protein. The idea that splicing factors may regulate one another to generate complex patterns of splicing is consistent with recent computational models⁴⁵. More generally, the role of complex splicing in the adult and developing nervous system is unclear. To answer the questions that come with increasingly complete transcriptomes in higher organisms, it will be necessary to study gene regulation downstream of transcription initiation, including the regulation of splicing, localization and translation.

METHODS SUMMARY

Animal staging, collection and RNA extraction. Tissues were dissected from Oregon R larval, pupal and adult staged animals synchronized with appropriate age indicators. Pupal and adult animals were treated with a number of environmental stresses. RNA was isolated using TRIzol (Invitrogen), treated with DNase and purified on a RNeasy column (Qiagen). Poly(A)⁺ RNA was prepared from an aliquot of each total RNA sample using an Oligotex kit (Qiagen).

RNA-seq. Libraries were generated and sequenced on an Illumina Genome Analyzer Ix or HiSeq 2000 using paired-end chemistry and 76-bp or 100-bp cycles. The 454 sequencing used poly(A)⁺ RNA from Oregon R adult males and females and mixed-staged *y¹ cn¹ bw¹ sp¹* embryos. Sequences are available from the Short Read Archive (Accession numbers available in Supplementary Table 10) and the modENCODE website (<http://www.modencode.org/>, Supplementary Table 10). CAGE⁴⁶ was sequenced on an Illumina Genome Analyzer Ix with 36-bp reads. Poly(A)⁺ seq was generated using a custom protocol (Supplementary Methods).

Analysis. RNA-seq, CAGE and poly(A)⁺ reads were mapped and filtered¹². GRIT was used to identify transcript models¹⁴. Expression levels for genes and exons were computed in BPKM⁶. GSC *P* values were computed⁴⁷. Ψ values were calculated with MISO⁴⁸. Differential expression analysis was conducted with a custom method (Supplementary Methods) and with DEseq⁴⁹. RPS-BLAST was used to conduct the conserved domain search with version v3.08 of the NCBI Conserved Domains Database (CDD) (Supplementary Methods). Orthology analysis between human and fly was conducted using DIOPT (http://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl). Phenotypic alleles were downloaded from FlyBase r5.50, and were selected as any allele localized to the genome with a disease phenotype.

Received 20 April; accepted 18 December 2013.

Published online 16 March 2014.

1. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
2. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
3. Takahashi, H., Kato, S., Murata, M. & Carninci, P. CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. *Methods Mol. Biol.* **786**, 181–200 (2012).
4. Mangone, M. *et al.* The landscape of *C. elegans* 3'UTRs. *Science* **329**, 432–435 (2010).
5. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* **469**, 97–101 (2011).
6. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
7. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2012).
8. Collins, J. E., White, S., Searle, S. M. & Stemple, D. L. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res.* **22**, 2067–2078 (2012).
9. Carninci, P. *et al.* Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res.* **13**, 1273–1289 (2003).
10. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
11. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
12. Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192 (2011).
13. Cherbas, L. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* **21**, 301–314 (2011).
14. Boley, N. *et al.* Genome guided transcript construction from integrative analysis of RNA sequence data. *Nature Biotechnol.* <http://dx.doi.org/10.1038/nbt.2850> (2014).
15. Celniker, S. E. & Rubin, G. M. The *Drosophila melanogaster* genome. *Annu. Rev. Genomics Hum. Genet.* **4**, 89–117 (2003).
16. Stapleton, M. *et al.* The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res.* **12**, 1294–1300 (2002).
17. Spradling, A. C. *et al.* The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**, 135–177 (1999).
18. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
19. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
20. Smibert, P. *et al.* Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Rep.* **1**, 277–289 (2012).
21. St Laurent, G. *et al.* Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in *Drosophila*. *Nature Struct. Mol. Biol.* **20**, 1333–1339 (2013).
22. Telonis-Scott, M., Kopp, A., Wayne, M. L., Nuzhdin, S. V. & McIntyre, L. M. Sex-specific splicing in *Drosophila*: widespread occurrence, tissue specificity and evolutionary conservation. *Genetics* **181**, 421–434 (2009).
23. Hartmann, B. *et al.* Distinct regulatory programs establish widespread sex-specific alternative splicing in *Drosophila melanogaster*. *RNA* **17**, 453–468 (2011).
24. Chang, P. L., Dunham, J. P., Nuzhdin, S. V. & Arbeitman, M. N. Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genomics* **12**, 364 (2011).
25. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
26. Lipshitz, H. D., Peattie, D. A. & Hogness, D. S. Novel transcripts from the *Ultrathorax* domain of the bithorax complex. *Genes Dev.* **1**, 307–322 (1987).
27. Tupry, J. L. *et al.* Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **102**, 5495–5500 (2005).
28. Young, R. S. *et al.* Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biol. Evol.* **4**, 427–442 (2012).
29. Kondo, T. *et al.* Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biol.* **9**, 660–665 (2007).
30. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
31. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
32. Duncan, D. M., Burgess, E. A. & Duncan, I. Control of distal antennal identity and tarsal development in *Drosophila* by *spineless-aristopedia*, a homolog of the mammalian dioxin receptor. *Genes Dev.* **12**, 1290–1303 (1998).
33. Schwartz, C., Locke, J., Nishida, C. & Kornberg, T. B. Analysis of *cubitus interruptus* regulation in *Drosophila* embryos and imaginal disks. *Development* **121**, 1625–1635 (1995).
34. Misra, S. *et al.* Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biology* **3**, research0083 (2002).
35. Lipovich, L. *et al.* Activity-dependent human brain coding/noncoding gene regulatory networks. *Genetics* **192**, 1133–1148 (2012).
36. Okamura, K. & Lai, E. C. Endogenous small interfering RNAs in animals. *Nature Rev. Mol. Cell Biol.* **9**, 673–678 (2008).
37. Okamura, K., Balla, S., Martin, R., Liu, N. & Lai, E. C. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nature Struct. Mol. Biol.* **15**, 581–590 (2008).
38. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**, 798–802 (2008).
39. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**, 1077–1081 (2008).
40. Engström, P. G. *et al.* Complex loci in human and mouse genomes. *PLoS Genet.* **2**, e47 (2006).
41. Whitehead, A. *et al.* Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes. *Proc. Natl Acad. Sci. USA* **109**, 20298–20302 (2012).
42. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
43. Bánfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).
44. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
45. Huelga, S. C. *et al.* Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).
46. Takahashi, H., Lassmann, T., Murata, M. & Carninci, P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nature Protocols* **7**, 542–561 (2012).
47. Bickel, P. J., Boley, N., Brown, J. B., Huang, H. & Zhang, N. R. Subsampling methods for genomic inference. *Ann. Appl. Stat.* **4**, 1660–1697 (2010).
48. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**, 1009–1015 (2010).
49. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
50. Yepiskoposyan, H. *et al.* Transcriptome response to heavy metal stress in *Drosophila* reveals a new zinc transporter that confers resistance to zinc. *Nucleic Acids Res.* **34**, 4866–4877 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the modENCODE transcription consortium, especially J. Landolin and J. Sandler for their early contributions to these studies. We also thank A. Kundaje and H. Huang for helpful discussions. This work was funded by a contract from the National Human Genome Research Institute modENCODE Project, contract U01 HG004271 and U54 HG006944, to S.E.C. (principal investigator) and P.C., T.R.G., R.A.H. and B.R.G. (co-principal investigators) with additional support from R01 GM076655 (S.E.C.) both under Department of Energy contract no. DE-AC02-05CH11231. J.B.B.'s work was supported by NHGRI K99 HG006698. Work in P.J.B.'s group was supported by the modENCODE DAC sub-award 5710003102, U01HG007031-01 and the ENCODE DAC 5U01HG004695-04. Work in Bloomington was supported in part by the Indiana METACyt Initiative of Indiana University, funded by an award from the Lilly Endowment. Work in E.C.L.'s group was supported by U01-HG004261 and RC2-HG005639.

Author Contributions J.A., T.R.G., B.R.G., R.A.H., T.C.K. and S.E.C. designed the project. J.A., P.Ch., T.R.G., B.R.G., R.A.H., J.B.B., B.O. and S.E.C. managed the project. R.E. designed treatment protocols and prepared biological samples. T.C.K., J.A. and L.C. oversaw biological sample production. B.D.E., D.M. and J.R. prepared biological samples. D.Z. and B.E. prepared RNA samples. J.A. oversaw RNA sample production. G.E.M., S.O. and L.Y. prepared Illumina RNA-seq libraries. A.M.S. prepared CAGE libraries. P.Ca. oversaw production of CAGE libraries. C.A.D., G.E.M., S.O., L.Y., S.P. and K.H.W. performed Illumina sequencing. B.R.G. and S.E.C. managed Illumina sequencing production. R.A.H. conceived the poly(A)⁺ seq method. R.W. and R.A.H. developed the poly(A)⁺ seq protocol and produced the libraries. K.M. performed 454 sequencing. C.Y., S.P. and K.H.W. performed cDNA library screens and full-insert cDNA sequencing. S.E.C. oversaw cDNA production. E.F. and N.B. installed and administered computer infrastructure for data storage and analysis. J.B.B., N.B., M.H.S., M.O.D., B.W.B., D.S., J.W.C., S.S., J.W., A.A.S., N.P., E.C.L., P.J.B. and B.R.G. developed analysis methods. J.B.B., N.B., M.H.S., M.O.D., B.W.B., A.S.H., E.F., R.A.H., S.S., D.S., L.C., G.R., J.H., J.W., A.A.S., E.C.L., K.H.W., B.R.G. and S.E.C. analysed data. N.B., J.B.B., M.H.S., K.H.W. and S.E.C. generated annotations. D.S. and B.O. analysed species validation data. S.S., J.W. and E.C.L. analysed 3' UTR and antisense data. A.S.H., E.F. and S.E.C. analysed image data. M.H.S. analysed proteomics data. M.H.S., S.S., D.S., B.O., E.C.L., T.C.K., R.E., R.A.H. and P.Ch. contributed to the text. A.S.H. assisted with manuscript preparation. J.B.B., B.R.G. and S.E.C. wrote the paper with input from all authors. All authors discussed the results and commented on the manuscript.

Author Information Sequences are available from the Short Read Archive and the modENCODE website, a list of accession numbers is given in Supplementary Table 10. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B.B. (benbrown@berkeley@gmail.com), B.R.G. (graveley@neuron.uchc.edu) or S.E.C. (celniker@fruitfly.org).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Regulatory analysis of the *C. elegans* genome with spatiotemporal resolution

Carlos L. Araya¹, Trupti Kawli¹, Anshul Kundaje², Lixia Jiang¹, Beijing Wu¹, Dionne Vafeados³, Robert Terrell³, Peter Weissdepp³, Louis Gevartzman³, Daniel Mace³, Wei Niu⁴, Alan P. Boyle¹, Dan Xie¹, Lijia Ma⁵, John I. Murray⁶, Valerie Reinke⁴, Robert H. Waterston³ & Michael Snyder¹

Discovering the structure and dynamics of transcriptional regulatory events in the genome with cellular and temporal resolution is crucial to understanding the regulatory underpinnings of development and disease. We determined the genomic distribution of binding sites for 92 transcription factors and regulatory proteins across multiple stages of *Caenorhabditis elegans* development by performing 241 ChIP-seq (chromatin immunoprecipitation followed by sequencing) experiments. Integration of regulatory binding and cellular-resolution expression data produced a spatiotemporally resolved metazoan transcription factor binding map. Using this map, we explore developmental regulatory circuits that encode combinatorial logic at the levels of co-binding and co-expression of transcription factors, characterizing the genomic coverage and clustering of regulatory binding, the binding preferences of, and biological processes regulated by, transcription factors, the global transcription factor co-associations and genomic subdomains that suggest shared patterns of regulation, and identifying key transcription factors and transcription factor co-associations for fate specification of individual lineages and cell types.

In multicellular organisms, transcription factors bind at *cis*-regulatory elements in the genome to mediate diverse gene expression programs with exquisite spatiotemporal control^{1–3}. However, owing to the paucity of *in vivo* developmental stage transcription factor binding data and cellular transcription factor expression data, the integrated maps required to study transcriptional control of development with spatiotemporal resolution are lacking.

In this work, we analyse regulatory activity of a broad set of *C. elegans* transcription factors in one or more developmental stages. Exploiting recently developed methods^{4–6}, we integrate transcription factor binding data with an initial cellular-resolution map of transcription factor expression in the embryo. Our integrated analyses support the discovery of many key transcription factors and candidate transcription factor co-associations for fate specification, providing insights into the temporal and spatial dynamics of regulatory interactions in development.

Large-scale survey of regulatory binding

As part of the modENCODE consortium, we performed 241 ChIP-seq experiments to identify *in vivo* binding sites for 92 (10%) *C. elegans* transcription factors and regulatory proteins (collectively termed factors) in one or more stages of development or treatments (Fig. 1a and Supplementary Table 1). To identify factor binding from the approximately 5.1 billion raw reads, we developed a uniform processing pipeline (Extended Data Fig. 1a–e and Methods) that enables comparison of orthologous transcription factor properties⁷, such as sequence preferences (Extended Data Fig. 1f–h). Eight previously reported⁸ experiments failed to pass our quality-control checks and were thus removed from consideration.

We focused our analyses on embryonic and larval (L1–L4) stages, examining a total of 397,539 reproducible binding sites distributed across 33,833 binding regions in the genome. Collectively, factor binding (excluding RNA polymerases) is spread throughout 21.7% of the *C. elegans* genome (Fig. 1b), an upper-bound defined by ChIP-seq resolution⁹.

We estimate that—within our ChIP-seq resolution and sensitivity—we have identified approximately 90% of the regulatory binding regions (albeit not the majority of binding events; Extended Data Fig. 1i). Consistent with this estimate, we observe binding within 2 kilobases (kb) upstream of a transcription start site (TSS) for 91.3% of genes (Extended Data Fig. 1j)¹⁰.

HOT regions are dynamic in development

Previous studies^{8,11–13} have revealed regions in metazoan genomes with heavily clustered transcription factor binding, termed high-occupancy target (HOT) regions. The availability of multiple data sets across stages allowed us to examine the dynamics of HOT regions through development. We identified HOT and extreme-occupancy target (XOT) regions for each developmental stage, where significant enrichments (false discovery rate (FDR) <5% and <1%, respectively) in transcription factor binding sites are observed (Extended Data Fig. 2a–c). We found a total of 9,142 HOT regions (spanning 2,948 genomic regions) in at least one developmental stage, and 858 constitutive HOT regions occurring across all stages assayed (Fig. 1b, Supplementary Table 2). Constitutive HOT regions are enriched in promoters of genes with house-keeping functions (Extended Data Fig. 2d and Supplementary Table 3). However, most HOT regions are dynamic across development: 31–56% of HOT regions change between sequential stages and occupancy at larval L4-specific HOT regions increases as development progresses (Fig. 1c).

Across developmental stages, 77–85% of HOT regions occur within 2 kb upstream of an annotated TSS (Extended Data Fig. 2e). Furthermore, 88.8% and 88.7% of constitutive HOT regions occur in promoter or enhancer states in embryos and L3 larvae, respectively (Extended Data Fig. 2f, g). These results indicate that HOT regions reside at important regulatory locations (Fig. 1d, Extended Data Fig. 3a–c) in the genome and are dynamic during development.

¹Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁴Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA. ⁵Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. ⁶Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

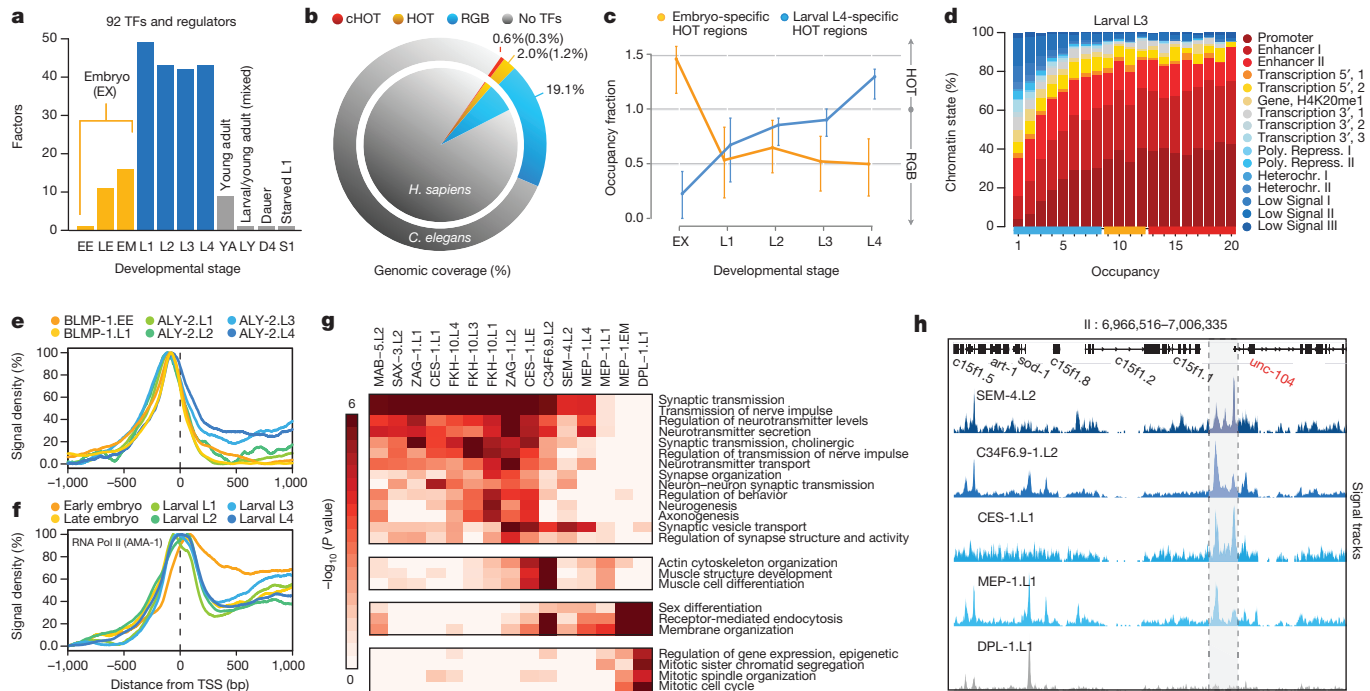


Figure 1 | Large-scale regulatory analysis of the *C. elegans* genome.

a, Transcription factors (TFs) and regulatory proteins assayed per developmental stage (or treatment) in 241 quality-filtered ChIP-seq experiments. Stages and treatments are early embryo (EE), late embryo (LE), embryo mixed (EM; EE and LE), larval L1 (L1), larval L2 (L2), larval L3 (L3), larval L4 (L4), young adult (YA), mixed larval and young adults (LY), day 4 adult (D4), and starved L1 (S1). Embryonic data sets were combined into a compiled embryonic stage (EX). Analyses in this report focus on embryonic (yellow) and larval (blue) experiments ($N = 187$). **b**, Genomic coverage (percent of genomic bases) of regulatory binding (excluding RNA polymerases) in 181 *C. elegans* (outer circle) and 339 *H. sapiens* (inner circle) ChIP-seq experiments. Genomic coverage of constitutive HOT (cHOT), HOT, and other regulatory binding (RGB) regions are highlighted in red, yellow and blue, respectively. Constitutive XOT (cXOT) and XOT percentages are shown in

parenthesis. cHOT, HOT and RGB region coverage in the human genome are 0.17%, 1.4% and 6.1%, respectively. **c**, Cut-off-normalized occupancy levels in 126 embryo-specific (yellow) and 91 larval L4-specific (blue) HOT regions. Bars indicate the 25th and 75th percentiles. **d**, Chromatin state, as determined in ref. 14, of L3 larvae binding regions by occupancy. RGB-, HOT- and XOT-region occupancy levels are indicated along the x axis as blue, yellow and red bars, respectively. Poly. Repress. and Heterochr. indicate Polycomb, repressed and heterochromatin states. **e, f**, Signal densities near enzymatically derived TSSs²⁹ for BLMP-1 and ALY-2, and RNA Pol II. **g**, Functional (GO term) enrichment for gene targets of binding³⁰. A subset of biological process terms (level ≥ 4) are shown for factors enriched (Benjamini–Hochberg-corrected, $P < 0.01$) in synaptic transmission; early MEP-1 and DPL-1 data sets are included for comparison. **h**, Example signal tracks near the UNC-104 locus.

Preferences and roles of regulators

Factors displayed a range of chromatin state¹⁴ preferences, with a general bias towards promoter and enhancer states (Extended Data Fig. 3d, e). Although generally clustered near TSSs, many factors display enrichments for upstream or downstream binding (Extended Data Fig. 3f). Proximal and downstream binders include RNA Pol II (AMA-1) and other regulators of transcription initiation and elongation, respectively. Upstream binders may be enriched for chromatin remodellers and factors that recruit the transcriptional machinery. For example, binding of BLMP-1—the orthologue of the human repressor PRDM1 (refs 15, 16)—is tightly concentrated upstream of TSSs (Fig. 1e). Likewise, ALY-2, a human THOC4 messenger RNA export factor orthologue¹⁷, exhibits increased binding downstream of TSSs during development (Fig. 1e) and is increased at elongation chromatin states relative to other factors. Generally, transcription factors assayed in multiple stages retain their upstream and downstream binding preferences. Remarkably, RNA Pol II positioning shifts (Fig. 1f) from a strong elongating distribution in the early embryo to weaker elongation distributions in later stages, consistent with its previously observed continued presence at promoters that are down-regulated during development^{8,18}.

Gene ontology (GO) analysis of the candidate protein-coding gene targets revealed 6,347 functional associations (BH-corrected, $P < 0.05$) for 75 factors (Extended Data Fig. 4a and Supplementary Table 4), suggesting biological roles for transcription factors of previously unknown function. The unstudied factors FKX-10 and C34F6.9 group with the established neuronal fate regulators SEM-4 (ZNF236), MAB-5

(HOXA2 and HOXB2), SAX-3 (ROBO1, ROBO2 and ROBO3), CES-1 and ZAG-1 in targeting neurotransmission genes, with C34F6.9 additionally regulating muscle development and sex differentiation (Fig. 1g). Most of these factors, including SEM-4 and C34F6.9 in L2 larvae, appear to regulate the neuronal kinesin UNC-104 (human KIF1A, KIF1C, Fig. 1h). Although expression of FKX-10 is restricted to six neuronal cells near the terminal bulb of the pharynx¹⁹, its specific molecular role in neuronal regulation and its regulatory targets were heretofore unknown.

Functional associations also demonstrate malleability of regulation. For example, UNC-62 transitions from regulating diverse muscle and neuronal development genes in embryos to regulating lipid metabolism processes in L4 larvae (Extended Data Fig. 4b). These changes are consistent with known diverse UNC-62 roles in motor neuron and vulval development, as well as locomotion, and ageing^{20,21}. Similarly, SAX-3 transitions from targeting neuronal genes in L2 larvae to targeting carbohydrate and lipid metabolism genes in L4 larvae. The change in UNC-62 regulatory targets coincides with increased expression of the UNC-62 (7a) isoform in late larval and adult intestine²⁰, which has been shown to affect lifespan²². Such early development regulators may often target metabolic regulation in later developmental stages²³.

Global and subdomain transcription factor co-associations

Global analyses of pairwise transcription factor co-associations²⁴ revealed a multitude of established and novel co-associations (Fig. 2a), many stage-specific clusters of co-association (Extended Data Fig. 5a), as well

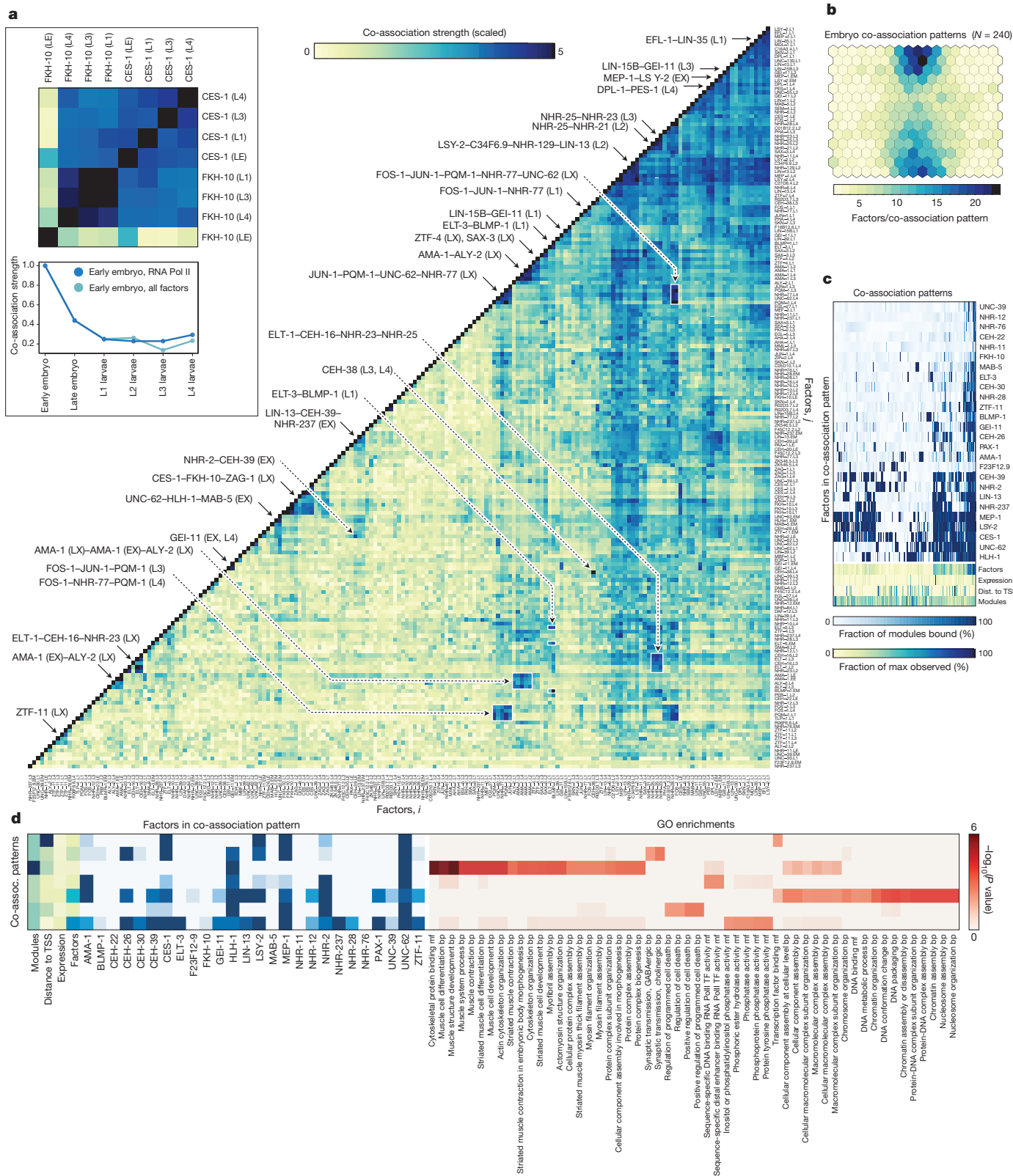


Figure 2 | Global and domain-specific patterns of transcription factor co-association. **a**, Matrix of global pairwise (i, j) factor co-association strengths ($N = 17,391$) as defined by promoter interval statistics²⁴. Co-association scores are scaled by the standard deviation (uncentered) for visualization purposes. Co-associations of interest and discussed in the text are highlighted. LX indicates larval stages L1–L4. A higher-resolution version is available in Extended Data Fig. 10. CES-1–FKH-10 co-associations are highlighted in the inset, top. Co-association strengths (unscaled) between early embryo and later stages are shown in the inset, bottom, for RNA Pol II-specific binding (blue), and for all factor-specific

binding (light blue). **b**, Embryonic (EX) binding regions ($N = 6,555$) were clustered into a SOM describing 240 co-association patterns among 26 factors. **c**, Binding signatures (fraction of modules bound by each factor) of the learned co-association patterns are shown. The relative number of factors per co-association pattern, expression from overlapping promoters, distance to TSSs, and number of modules with each co-association pattern are indicated as a fraction of the maximum observed across co-association patterns. **d**, Functional enrichment for regions with UNC-62-bound co-association patterns of the embryo SOM. Molecular function (mf) and biological process (bp) terms are shown.

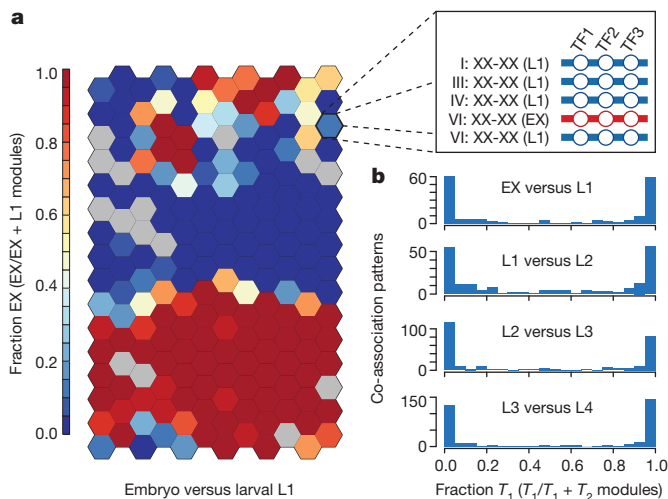


Figure 3 | Stage specificity in higher-order transcription factor co-associations. **a**, Embryonic (EX) and larval L1 binding SOM with raw binding sites. Binding data for factors ($N = 15$) assayed in embryos and L1 larvae was assigned to 25,261 stage-specific binding modules as shown in the inset. Stage-specific binding modules were clustered into a SOM describing 192 co-association patterns. The SOM is coloured by the embryonic (versus L1) stage specificity of the learned co-association patterns, measured as the fraction of binding modules that are embryonic. **b**, Histogram of preceding (T_1) versus subsequent (T_2) stage specificities.

as differences in co-associations between expressed and repressed promoters (Extended Data Fig. 5b). FOS-1-JUN-1 as well as GEI-11-LIN-15B co-associations are readily apparent in L1 and L3 larvae, but not in L4 larvae. Likewise, ELT-3 and BLMP-1, which preferentially reside at molting and cuticle development gene promoters, co-localize in L1 larvae but not in embryos. The neuronal regulators CES-1 and FKH-10 co-associate across larval stages (L1, L3–L4) but their co-association is not apparent in late embryogenesis (Fig. 2a). Changes in co-association are often correlated with the presence of additional factors, for example, in the embryo to larval L1 transition, the increased ELT-3-BLMP-1 co-association is also accompanied by increased GEI-11 co-associations with these factors (Extended Data Fig. 5c–f). Other factors remain largely invariant through multiple stages, for example, ZTF-11, a human MTF1 orthologue.

Functionally related factors were often co-associated. For example, FOS-1, NHR-77 and PQM-1 target promoters of genes in cellular lipid and ketone metabolic processes. Similarly, EFL-1 and LIN-35, the known interacting orthologues of human E2F and RB, show a strong co-association in L1 larvae, where they target membrane organization and endocytosis genes.

We observed strong similarities in RNA Pol II binding within embryonic (early and late embryo) and within larval L1–L4 stages, but larval RNA Pol II binding is only marginally and weakly co-associated with embryonic binding, reflecting the dynamic establishment of the transcriptional machinery through development (Fig. 2a).

To uncover higher-order co-associations (involving two or more factors), and their genomic subdomains, we applied self-organizing maps (SOMs), an unsupervised machine learning technique²⁵. For each developmental stage, we trained SOMs to cluster genomic regions with shared transcription factor co-association patterns (Fig. 2b and Extended Data Fig. 6a–d), thereby concomitantly identifying transcription factor co-association patterns (Fig. 2c) and their target regulatory regions.

We performed GO analysis of the target regulatory regions for 240, 390, 439, 390 and 409 clusters in the embryo, larval L1, L2, L3 and L4 SOMs, respectively, revealing enrichments across 1,209 GO terms (BH-corrected, $P < 0.05$, Extended Data Fig. 6e and Supplementary Table 5). As illustrated in the embryo, higher-order co-association patterns show a richness of functional associations, with 137 clusters spanning 273 GO

terms. A close examination of UNC-62 co-association patterns reveals how diverse patterns for individual factors can result in specialized functional targeting (Fig. 2d). Regions bound exclusively by UNC-62 and HLH-1 are highly-enriched at muscle development promoters. In contrast, genes targeted by more complex UNC-62 co-associations are enriched in synaptic transmission, regulation of cell death, and chromatin assembly functions. Higher-order co-associations are largely stage-specific (Fig. 3), a feature modulated by changes in the observed number of binding sites for individual factors between stages (Extended Data Fig. 7).

Spatiotemporal transcription factor expression analysis

Although studies in *C. elegans* and *D. melanogaster* have led analyses of organismal-level regulatory binding circuits, such studies have generally lacked cell-type and tissue resolution. We sought to remedy this deficiency by tracking^{5,6} the expression of 180 diverse genes (mostly transcription factors) through early embryogenesis with cellular resolution (Extended Data Fig. 8a–d). Our expression data, from previously published^{5,6} and newly acquired series, includes 36 factors with genome-wide binding measurements (13 embryo, 23 larval).

We observed common and distinctive cellular expression patterns amongst a wide distribution of broadly- and narrowly-expressed genes (Extended Data Fig. 8e, f). For example, expression of DMD-4, an orthologue of the vertebrate spinal circuit configuration regulator DMRT3, is tightly limited to posterior regions of the pharynx. Similarly, F49E8.2 expression is exclusive to the Z2 and Z3 germ cells (Extended Data Fig. 8a). 95.7% of pairs of tracked cells show distinct gene expression signatures ($R < 0.75$).

Cellular expression mapped the regulatory activity of 16 assayed factors to specific tissues (Fig. 4a). As expected, the known regulators of pharynx and muscle development, PHA-4 and HLH-1, were respectively enriched in these tissues. The co-associated factors, MEP-1 and DPL-1 (human DP1 and DP2 orthologue), although broadly expressed, are enriched in neuronal lineages. This is consistent with the observed MEP-1 targeting of neuronal function genes in the larvae, and provides further support for the coordinate activities of MEP-1 and DPL-1 in targeting membrane organization, receptor-mediated endocytosis, and cell-cycle genes (Fig. 1f and Supplementary Table 4).

More complex patterns of co-expression and co-association were observed in epidermal tissues, where CEH-16 (human Engrailed), and particularly ELT-1 and NHR-25 expression is concentrated. In both L2 and L3 larvae, ELT-1 and NHR-25 are modestly co-associated. ELT-1 targets transcriptional regulators, including NHR-25, and tail morphogenesis genes, whereas NHR-25 targets nuclear organization and genitalia development genes (Supplementary Table 4). However, larval L2 binding of ELT-1 and NHR-25 is co-associated with that of CEH-16, whose early embryonic expression is primarily concentrated in a subset of pharynx and epidermal cells.

In early embryo (Fig. 4a) and L1 larvae²⁶, HLH-1, is primarily and broadly expressed in muscle tissues whereas posterior-specific HOX factors MAB-5, and EGL-5 are expressed in a small subset of posteriorly placed muscle, epidermal and neuronal precursors. We observed modest co-association signals between embryonic HLH-1 and MAB-5 binding, and larval L3 EGL-5 binding, perhaps reflecting the intersection of tissue-specific and positional regulatory programs. As expected, HLH-1 targets muscle differentiation genes (together with UNC-62); however in GO analysis, we only detect MAB-5 targeting of diverse neuronal functions (in mixed embryos and L2 larvae), consistent with its later role in neuron specification²⁷. CO5D10.1, whose early embryonic expression is also restricted to muscle tissues is not co-associated with the above factors, and neither CO5D10.1 nor EGL-5 showed specific functional associations. Thus, although co-associated factors were often expressed in the same tissue, this is not pervasive. Moreover, these co-expression patterns are dynamically established during embryogenesis (data not shown).

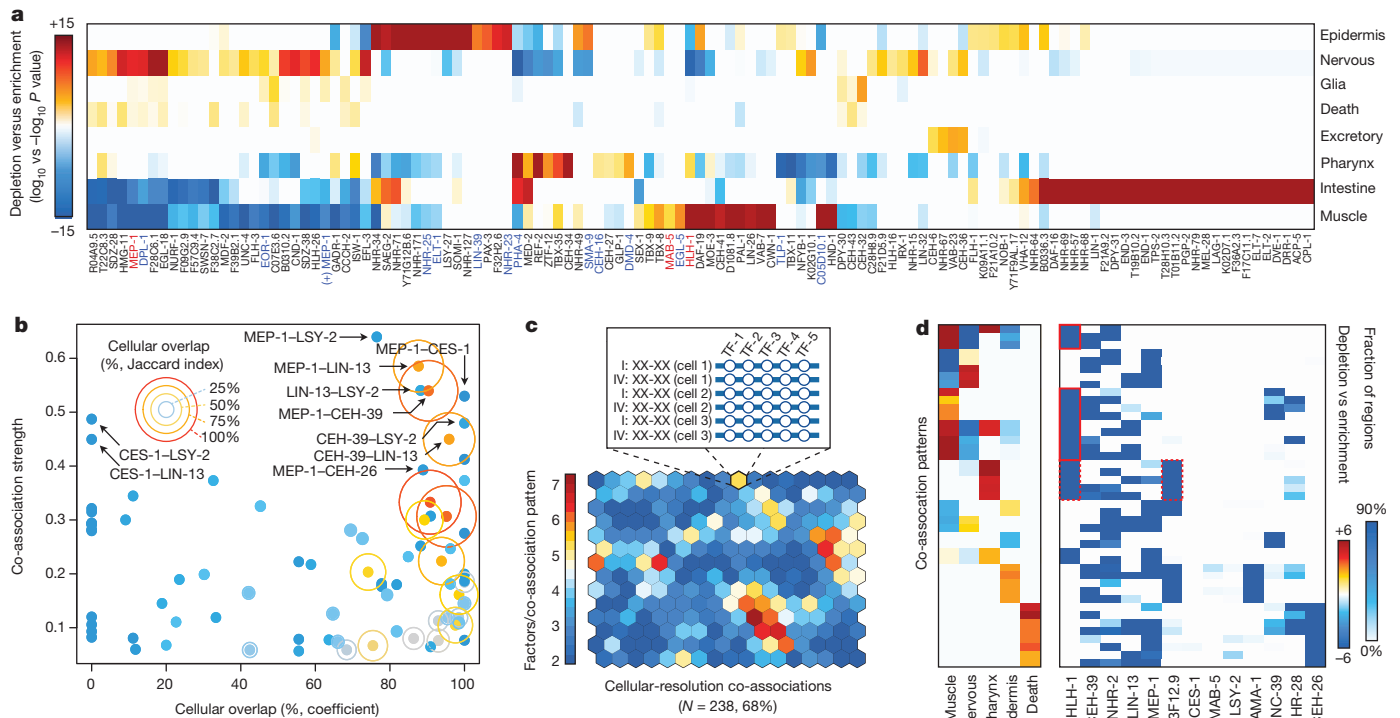


Figure 4 | Cell-type and lineage resolution of regulator activity and transcription factor co-associations. **a**, Tissue enrichment ($-\log_{10}$ P value) and depletion (\log_{10} P value) scores for the expressing population of each gene are shown (Fisher's exact, Bonferroni-corrected). Only genes with significant enrichments (or depletions) are shown. **b**, Co-association strength (Fig. 2a) versus cellular overlap coefficient for 13 focus factors. The Jaccard index for the cellular overlap is indicated for each gene pair by ring size and colour. **c**, Cellular-resolution regulatory binding SOM. Cellular-resolution binding modules were generated by annotating in each cell, the binding of focus factors

expressed in the cell. Cellular-resolution binding modules (inset) were clustered into a SOM with 268 learned co-association patterns, 161 (68%) of which were discovered in the data. The SOM is colored by the number of factors in the learned co-association patterns. **d**, Tissue classes and co-association signatures are shown for 43 co-association patterns with significant enrichments. Tissue enrichments of interest are highlighted red. Circle size and colour indicate the Jaccard index (%) of cellular overlap. 'Death' and 'Excretory' represent tissue designations for cells with apoptotic and excretory cell fates.

Refinement of embryonic co-associations

Despite extensive studies in metazoan regulatory networks, the relationship between regulator binding in overlapping genomic regions and co-expression in cell-types is not well studied. We examined this relationship among 13 'focus' factors, for which both embryonic binding and cellular expression were assayed. This analysis is limited to the first half of embryogenesis, where expression was directly measured in 696 'focus' cells. Later events may occur that would not be identified in our analysis. We found a poor correlation between transcription factor co-expression and co-association ($R = 0.07$, Fig. 4b), consistent with precise coordination of these separate processes underlying the differential establishment of cell- and lineage-specific regulatory circuits.

Integrated analysis shows that MEP-1 is co-associated and co-expressed with similarly broadly expressed factors (LIN-13, CEH-39) and narrowly expressed factors (CES-1, CEH-26), suggesting that MEP-1 often works in *cis* with these additional factors. MEP-1 binding is co-associated with CES-1 and CEH-26 in embryos, and expression of these factors is narrowly restricted within the MEP-1-expressing population. These MEP-1-CES-1 and MEP-1-CEH-26 co-associations are reminiscent of MEC-3-UNC-86 interactions in which the classic 'terminal selector' MEC-3 heterodimerizes with the broadly expressed UNC-86 exclusively in touch sensory neurons²⁸. Thus, the co-association and co-expression of MEP-1-CEH-26 suggests CEH-26 may function as a terminal selector in head and tail neurons, and the excretory cell. The spatiotemporally-resolved co-association analyses demonstrate how broadly expressed factors, such as MEP-1 and LIN-13—which targets both neurotransmission functions and genitalia development—can have diversified functional roles during development through co-associations with narrowly expressed factors.

To determine how co-binding and co-expression co-ordinately define regulatory patterning in distinct cell-types and genomic regions, we intersected cellular expression and binding data by mapping focus factor binding to *in silico* genomes for cells where the factors are expressed. This procedure resulted in 4,779,810 binding sites distributed across 2,858,477 cell-resolved binding modules. We applied an SOM to cluster the cell-resolved binding modules by co-association patterns, uncovering 161 transcription factor co-association patterns and the genomic subdomains and specific cellular subsets of the embryo in which they may occur (Fig. 4c). The cellular distribution of transcription factor co-association patterns revealed co-associations shared among and unique to specific cell fates (Fig. 4d). For example, we found that distinct MEP-1, CEH-26 and NHR-2 co-associations were specific to neuronal tissues. Similarly, muscle cells were enriched in various HLH-1 co-associations.

We identified 39 co-association patterns whose cellular distribution coincides with the cellular expression of at least one of 124 target genes (non-focus factors; Bonferroni-corrected, $P < 0.01$). Focus factor binding allowed us to analyse co-association patterns at the promoters of 44 of these genes (where binding is observed). For 28 (63.6%) of these genes, co-association patterns were detected at the promoter and the gene's cellular expression matched the cellular distribution. Moreover, the overlap between the expression cells for a gene and the co-association cells is higher in cases where the co-association occurs in the promoter of the gene (Wilcoxon, $P = 5.1 \times 10^{-6}$, Extended Data Fig. 8g). This result indicates that co-associations at promoters are correlated with cellular expression patterns for genes, and suggests a functional regulatory role for the discovered co-associations.

Discussion

We have generated a high-coverage transcription factor binding map of the *C. elegans* genome, revealing regulatory targets, co-associations, and dynamics across five developmental windows for 92 diverse factors. Gene targets suggest a multitude of functional associations for 75 factors, many previously unannotated and with clear mammalian homologues. Our work reveals extensive regulatory rewiring through development, with temporal differentiation of HOT regions in the genome, factor positioning preferences, regulatory targets, and co-associations.

A systematic analysis of transcription factor co-associations through development reveals sets of factors that assemble at genomic regions associated with more than 1,200 biological functions (GO terms), with probable spatiotemporal specificity. As illustrated with UNC-62, these higher-order co-associations reveal how individual transcription factors can participate in distinct transcription factor co-associations patterns at promoters of functionally diverse genes.

Lastly, cellular-resolution expression tracking allowed us to map the activity of 35 factors to precise cell and tissue types, demonstrating lineage-specific activities for 16 factors in the early embryo. Importantly, co-associations that are observed in whole-organism binding data are not always evident at the cellular level, highlighting the need to incorporate such information in our understanding of regulatory circuits. As additional expression patterns and transcription factor binding sites are determined, and methods to track transcription factor binding with cell-type specificity are developed, the broader and more precise regulatory logic of development should emerge.

METHODS SUMMARY

ChIP-seq assays of wild-type (N2) and transgenic nematodes were performed under controlled conditions (Extended Data Fig. 9) as described²³. Experimental and computational methods are described in detail in the Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 June 2013; accepted 22 May 2014.

- Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).
- Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Rev. Genet.* **13**, 613–626 (2012).
- Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
- Bao, Z. *et al.* Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **103**, 2707–2712 (2006).
- Murray, J. I. *et al.* Multidimensional regulation of gene expression in the *C. elegans* embryo. *Genome Res.* **22**, 1282–1294 (2012).
- Mace, D. L., Weisdepp, P., Gevirtzman, L., Boyle, T. & Waterston, R. H. A high-fidelity cell lineage tracing method for obtaining systematic spatiotemporal gene expression patterns in *Caenorhabditis elegans*. *G3 (Bethesda)* **3**, 851–863 (2013).
- Boyle, A. P. *et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* <http://dx.doi.org/10.1038/nature13668> (this issue).
- Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
- Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
- Allen, M. A., Hillier, L. W., Waterston, R. H. & Blumenthal, T. A global analysis of *C. elegans* trans-splicing. *Genome Res.* **21**, 255–264 (2011).
- Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
- Nègre, N. *et al.* A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
- Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13**, R48 (2012).

- Ho, J. W. K. *et al.* Comparative analysis of metazoan chromatin organization. *Nature* <http://dx.doi.org/10.1038/nature13415> (this issue).
- Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207–213 (2005).
- Smith, M. A. *et al.* PRDM1/Blimp-1 controls effector cytokine production in human NK cells. *J. Immunol.* **185**, 6058–6067 (2010).
- Kuersten, S., Segal, S. P., Verheyden, J., LaMartina, S. M. & Goodwin, E. B. NXF-2, REF-1, and REF-2 affect the choice of nuclear export pathway for *tra-2* mRNA in *C. elegans*. *Mol. Cell* **14**, 599–610 (2004).
- Baugh, L. R., Demodena, J. & Sternberg, P. W. RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324**, 92–94 (2009).
- Hope, I. A., Mounsey, A., Bauer, P. & Aslam, S. The forkhead gene family of *Caenorhabditis elegans*. *Gene* **304**, 43–55 (2003).
- Van Aiken, K. *et al.* Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in *C. elegans* embryogenesis. *Development* **129**, 5255–5268 (2002).
- Curran, S. P. & Ruvkun, G. Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS Genet.* **3**, e56 (2007).
- Van Nostrand, E. L., Sánchez-Blanco, A., Wu, B., Nguyen, A. & Kim, S. K. Roles of the developmental regulator unc-62/Homothorax in limiting longevity in *Caenorhabditis elegans*. *PLoS Genet.* **9**, e1003325 (2013).
- Zhong, M. *et al.* Genome-wide identification of binding sites defines distinct functions for *Caenorhabditis elegans* PHA-4/FOXA in development and environmental response. *PLoS Genet.* **6**, e1000848 (2010).
- Chikina, M. D. & Troyanskaya, O. G. An effective statistical evaluation of ChIP-seq dataset similarity. *Bioinformatics* **28**, 607–613 (2012).
- Xie, D. *et al.* Dynamic trans-acting factor colocalization in human cells. *Cell* **155**, 713–724 (2013).
- Liu, X. *et al.* Analysis of cell fate from single-cell gene expression profiles in *C. elegans*. *Cell* **139**, 623–633 (2009).
- Salser, S. J. & Kenyon, C. A. *C. elegans* Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. *Development* **122**, 1651–1661 (1996).
- Hobert, O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl Acad. Sci. USA* **105**, 20067–20071 (2008).
- Gu, W. *et al.* CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* **151**, 1488–1500 (2012).
- Zhu, L. J., Gazin, C. & Green, M. R. ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237–246 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Euskirchen at the Stanford Center for Genomics and Personalized Medicine for sequencing ChIP libraries and members of the Reinke laboratory for contributing ChIP-seq data. We thank members of the Waterston laboratory, Sarov laboratory and Kim laboratory for tagged constructs and generating *C. elegans* strains. We thank J. Reuter, D. Phanstiel, H. Tilgner and D. Fowler for critical comments on the manuscript. This work is supported by the NHGRI as part of the modENCODE project (U01 HG004267).

Author Contributions C.L.A., T.K., J.I.M., V.R., R.H.W. and M.P.S. designed experiments. T.K., L.J., L.M., W.N. and B.W. constructed strains and prepared ChIP libraries for sequencing. J.I.M., D.V., R.T., P.W., L.G. and D.M. constructed strains and performed reporter imaging experiments. A.K., C.L.A. and A.P.B. performed alignments, uniform binding site identification and quality filtering. C.L.A. organized data, pre-processed binding site and expression data, and executed all analyses. A.P.B., D.X. and C.L.A. developed SOM analysis techniques. C.L.A., V.R., J.I.M., R.H.W. and M.P.S. wrote the manuscript.

Author Information ChIP-seq data reported here are available through the modENCODE DCC (<http://www.modencode.org>) and at <http://tapanti.stanford.edu/cetrn>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P.S. (mepsnyder@stanford.edu) and R.H.W. (watersto@u.washington.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Cobalt-56 γ -ray emission lines from the type Ia supernova 2014J

E. Churazov^{1,2}, R. Sunyaev^{1,2}, J. Isern³, J. Knödseder^{4,5}, P. Jean^{4,5}, F. Lebrun⁶, N. Chugai⁷, S. Grebenev¹, E. Bravo⁸, S. Sazonov^{1,9} & M. Renaud¹⁰

A type Ia supernova is thought to be a thermonuclear explosion of either a single carbon–oxygen white dwarf or a pair of merging white dwarfs. The explosion fuses a large amount of radioactive ^{56}Ni (refs 1–3). After the explosion, the decay chain from ^{56}Ni to ^{56}Co to ^{56}Fe generates γ -ray photons, which are reprocessed in the expanding ejecta and give rise to powerful optical emission. Here we report the detection of ^{56}Co lines at energies of 847 and 1,238 kiloelectronvolts and a γ -ray continuum in the 200–400 kiloelectronvolt band from the type Ia supernova 2014J in the nearby galaxy M82. The line fluxes suggest that about 0.6 ± 0.1 solar masses of radioactive ^{56}Ni were synthesized during the explosion. The line broadening gives a characteristic mass-weighted ejecta expansion velocity of $10,000 \pm 3,000$ kilometres per second. The observed γ -ray properties are in broad agreement with the canonical model of an explosion of a white dwarf just massive enough to be unstable to gravitational collapse, but do not exclude merger scenarios that fuse comparable amounts of ^{56}Ni .

The detailed physics of the explosion of type Ia supernovae (for example deflagration or detonation) and the evolution^{4,5} of a compact object towards explosion remain a matter of debate^{6–9}. In a majority of models, the ejecta are opaque to γ -ray lines during first 10–20 days after the explosion (because of Compton scattering). At later times, the ejecta become progressively more transparent and a large fraction of γ -rays escapes. This leads to a robust prediction¹⁰ of γ -ray emission from type Ia supernovae after few tens of days, dominated by the γ -ray lines of ^{56}Co . Such emission has been observed before: the down-scattered hard X-ray continuum from supernova (SN) 1987A in the Large Magellanic Cloud was seen half a year after the explosion^{11,12}, and γ -ray lines of ^{56}Co were detected several months later^{13,14}. That was a core-collapse (type II) supernova, in which the cause of the explosion is completely different from that of type Ia supernovae. Type Ia events, despite being intrinsically brighter, are more rare than core-collapse supernovae, and before SN 2014J there was not one close enough to detect. The recent type Ia SN 2011fe at a distance of $D \approx 6.4$ Mpc yielded only an upper limit on the ^{56}Co line flux¹⁵.

SN 2014J in M82 was discovered¹⁶ on January 21, 2014. The reconstructed¹⁷ date of the explosion is January 14.75 UT. This is the nearest type Ia supernova to be detected in at least four decades, at the M82 distance of $D \approx 3.5$ Mpc (ref. 18). The European Space Agency satellite INTEGRAL¹⁹ started observing SN 2014J in 2014, from January 31 to April 24 and from May 27 to June 26. We use the INTEGRAL data taken between days 50 and 100 after the explosion, the period when the expected flux from γ -ray lines of ^{56}Co is close to the maximum¹⁰. This set of observations by the SPI and ISGRI/IBIS instruments on board INTEGRAL has been analysed, excluding periods of strong solar flares, which cause large variations in the instrumental background (Methods and Extended Data Fig. 1). The spectrum derived assuming a point source at the position of SN 2014J is shown in Fig. 1 using red and blue points for SPI and ISGRI, respectively.

The model spectrum is binned similarly to the observed supernova spectrum. The signatures of the 847 and 1,238 keV lines are clearly seen in the spectrum (along with tracers of weaker lines of ^{56}Co at 511 and 1,038 keV). The low-energy (<400 keV) part of the SPI spectrum is not shown because of possible contamination due to off-diagonal response of the instrument to higher-energy lines. At these energies, we use ISGRI/IBIS data instead (Methods).

By varying the assumed position of the source and repeating the flux-fitting procedure using SPI data (Methods) we construct a $40^\circ \times 40^\circ$ image of the signal-to-noise ratio in the 800–880 and 1,200–1,300 keV energy bands (Fig. 2). SN 2014J is detected at 3.9 s.d. and 4.3 s.d. in these two bands, respectively. These are the highest peaks in both images.

The images obtained by ISGRI at lower energies (100–600 keV) during the observations of SN 2014J and in October–December 2013, that is, a few months before the SN 2014J explosion (see Methods for the details of the earlier observation), are shown in Fig. 3. An inspection of images in the 25–50 keV band shows that the fluxes observed in 2013 and 2014 are similar, whereas at higher energies (>100 keV) there is excess at the position of SN 2014J only in the 2014 data. Previous ISGRI observations of this field in 2009–2012, with a total exposure of about 6×10^6 s, revealed no significant signal at energies >50 keV from M82²⁰.

A combination of imaging and spectral analysis provides robust evidence of γ -ray emission from SN 2014J. As expected, much of the signal comes from the 800–900 and 1,200–1,300 keV bands, where two

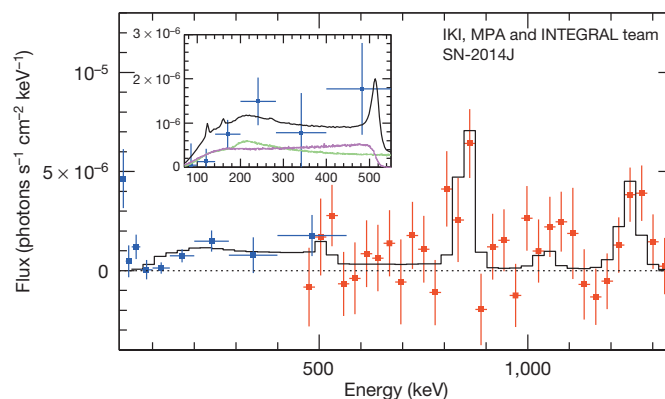


Figure 1 | Gamma-ray lines from Co decay at 847 and 1,238 keV in the spectrum of SN 2014J. The spectrum was obtained by INTEGRAL between days 50 and 100 after the outburst. Red and blue points show SPI and ISGRI/IBIS data, respectively. The flux below 60 keV is dominated by the emission of M82. The black curve shows a fiducial model of the supernova spectrum for day 75 after the explosion. Inset, lower-energy part of the spectrum (black). The expected contributions of three-photon positronium annihilation (magenta) and Compton down-scattered emission from 847 and 1,238 keV lines (green) are also shown. All error bars, 1 s.d.

¹Space Research Institute (IKI), Profsoyuznaya 84/32, Moscow 117997, Russia. ²Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, 85741 Garching, Germany. ³Institute for Space Sciences (ICE-CSIC/IEEC), 08193 Bellaterra, Spain. ⁴Université de Toulouse, UPS-OMP, IRAP, Toulouse, France. ⁵CNRS, IRAP, 9 Avenue colonel Roche, BP 44346, F-31028 Toulouse Cedex 4, France. ⁶APC, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 75205 Paris Cedex 13, France. ⁷Institute of Astronomy of the Russian Academy of Sciences, 48 Pyatnitskaya Street, 119017 Moscow, Russia. ⁸ETSAV, Universitat Politècnica de Catalunya, Carrer Pere Serra 1-15, 08173 Sant Cugat del Valles, Spain. ⁹Moscow Institute of Physics and Technology, Institutskiy pereulok 9, 141700 Dolgoprudny, Russia. ¹⁰LUPM, Université Montpellier 2, CNRS/IN2P3, CC 72, Place Eugène Bataillon, F-34095 Montpellier Cedex 5, France.

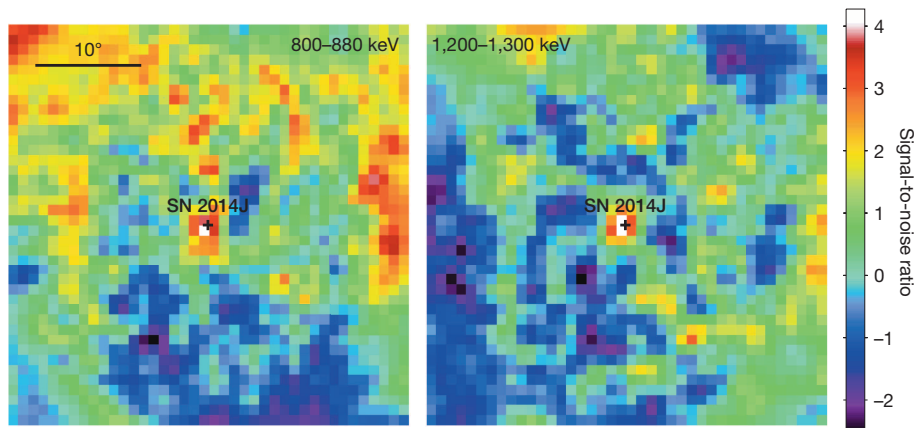


Figure 2 | Signatures of ^{56}Co lines at 847 and 1,238 keV in SPI images. The broad energy bands 800–880 keV and 1,200–1,300 keV are expected to contain the flux from ^{56}Co lines, accounting for the expected broadening and shifting due to the ejecta expansion and opacity effects. The source is detected at 3.9 s.d. and 4.3 s.d., respectively, in these two bands. Colours show the signal-to-noise ratio at a given position. Data courtesy of IKI, MPA and the INTEGRAL team.

prominent lines of ^{56}Co should be. The best-fitting parameters (flux, energy and broadening) of those two lines are given in Extended Data Table 1. The fluxes at 847 and 1,238 keV, respectively $(2.34 \pm 0.74) \times 10^{-4}$ and $(2.78 \pm 0.74) \times 10^{-4}$ photons $\text{s}^{-1} \text{cm}^{-2}$, were corrected for the fraction of γ -ray photons escaping the ejecta (this fraction is $\sim 60\%$; Methods), yielding an estimate of $M_{\text{Co}} = (0.34 \pm 0.07)M_{\odot}$ for the total ^{56}Co mass at day 75 (M_{\odot} , solar mass). Correcting this value for the secular evolution of the ^{56}Co mass in the decay chain $^{56}\text{Ni} \rightarrow ^{56}\text{Co} \rightarrow ^{56}\text{Fe}$, provides an estimate of the initial ^{56}Ni mass: $M_{\text{Ni}} = (0.61 \pm 0.13)M_{\odot}$.

An independent estimate of the ^{56}Ni mass can be obtained from the optical ‘bolometric’ light curve. The simplest approach is based on the assumption²¹ that the bolometric luminosity at the maximum is approximately equal to the power of the radioactive decay at this moment. For SN 2014J, the maximum bolometric luminosity is $\sim 1.1 \times 10^{43} \text{ erg s}^{-1}$, attained on day 17.7 after the explosion²², assuming interstellar extinction $A_V = 1.7$ mag in the V filter band. A thorough analysis of extinction²³ yields $A_V = 1.85 \pm 0.11$ mag. This implies a ^{56}Ni mass of $(0.42 \pm 0.05)M_{\odot}$. This is marginally consistent with the estimates based on the γ -ray emission lines, which is not surprising, given the qualitative nature of this estimate and the large and complicated extinction in the direction of SN 2014J. A more direct test is the comparison of γ -ray and bolometric optical, ultraviolet and infrared luminosities at day 75. The latter is estimated to be $\sim 1.1 \times 10^{42} \text{ erg s}^{-1}$ (Methods). The total energy released during decay of the ^{56}Co isotope²⁴ is split between neutrinos (~ 0.8 MeV), kinetic energy of positrons (~ 0.12 MeV) and γ -rays (~ 3.6 MeV). In our fiducial model a fraction $f \approx 0.77$ of the luminosity in γ -rays escapes the ejecta at day 75. The remaining fraction, $1 - f \approx 0.23$, is deposited in the ejecta (ignoring bremsstrahlung radiation losses by electrons). Adding the kinetic energy of positrons and 23% of the γ -ray luminosity produced by $0.34M_{\odot}$ of ^{56}Co yields an estimate of $\sim 1.1 \times 10^{42} \text{ erg s}^{-1}$ for the rate of energy deposition in the ejecta, in good agreement with the optical data. The same model predicts that $\sim 3.3 \times 10^{42} \text{ erg s}^{-1}$ escape the ejecta in the form of hard X-rays and γ -rays. The observed luminosities of the 847 and 1,238 keV lines are $\sim 4.7 \times 10^{41}$ and $\sim 8.1 \times 10^{41} \text{ erg s}^{-1}$, respectively.

The emergent lines are expected to be broadened and blueshifted because of ejecta expansion and the opacity effects (Methods and Extended Data Fig. 3). Both effects are indeed observed (Fig. 4). The mean blueshift, averaged over both lines, corresponds to a velocity of $V_{\text{shift}} = -3,100 \pm 1,100 \text{ km s}^{-1}$, and the broadening (root mean squared line-of-sight velocity) is $\sigma_{\gamma} = 4,100 \pm 960 \text{ km s}^{-1}$. These values are broadly consistent with expectations of our fiducial model (Methods).

Finally, at lower energies (100–400 keV) the emerging flux is dominated by Compton scattering of the 847 and 1,238 keV photons and the ortho-positronium continuum from positron annihilation (Fig. 1 inset and Extended Data Fig. 4). The predicted 100–400 keV flux is consistent with INTEGRAL data. Roughly half of the signal comes from the ortho-positronium continuum, suggesting that at day 75 the positrons produced in 19% of the ^{56}Co decay are thermalized in the ejecta and annihilate via positronium formation. Below 100 keV, the emission declines strongly owing to photoabsorption.

It should be possible to derive stronger constraints from the combined analysis of the whole spectrum. As a first step, we constructed a three-parameter (M_{Ni} ; ejecta mass, M_{Ej} ; characteristic velocity, V_e) model that is capable of reproducing the main observables in the spectrum: the γ -ray line flux, the line broadening and the continuum flux below 511 keV (Methods). The model assumes spherical symmetry, complete mixing of all elements over the entire ejecta, and an exponential density profile^{25,26}: $\rho \propto e^{-r/V_e}$. All three parameters are treated as independent. The level of mixing can in principle be determined from the time evolution of the γ -ray flux. For example, early appearance of hard X-ray emission from core-collapsed SN 1987A clearly demonstrated that Co is mixed²⁷ over the ejecta. A Monte Carlo code follows the propagation of the γ -ray photons through the ejecta and accounts for scattering and photoabsorption of photons and annihilation of positrons (Methods).

We vary M_{Ni} , M_{Ej} and V_e over a wide range, calculate the expected emergent spectrum and compare it (in terms of a χ^2 test) with the observed spectrum. This procedure yields the following best-fit parameters and 1 s.d. confidence intervals for the individual parameters: $M_{\text{Ni}} = 0.56^{+0.14}_{-0.06}M_{\odot}$, $M_{\text{Ej}} = 1.2^{+1.9}_{-0.5}M_{\odot}$, $V_e = 3,000 \pm 800 \text{ km s}^{-1}$. In this

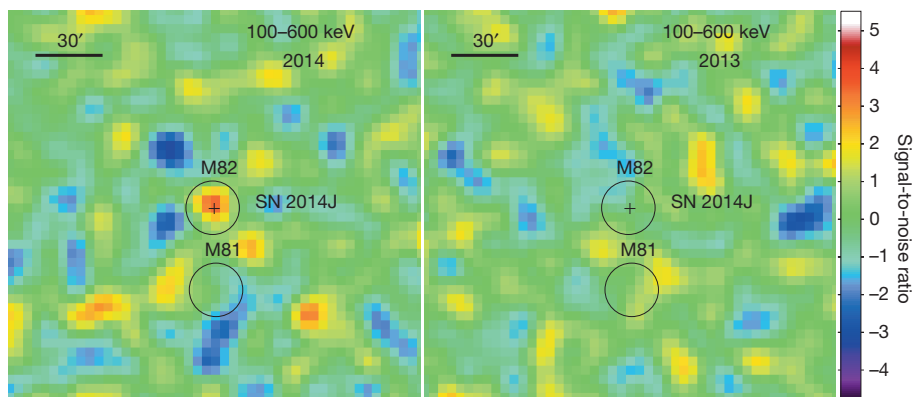


Figure 3 | Appearance of a new hard (100–600 keV) X-ray source at the position of SN 2014J. In the ISGRI image of the M82 field taken in 2013 the source is absent. Colours show the signal-to-noise ratio at a given position. SN 2014J is detected in this image at ~ 3.7 s.d.

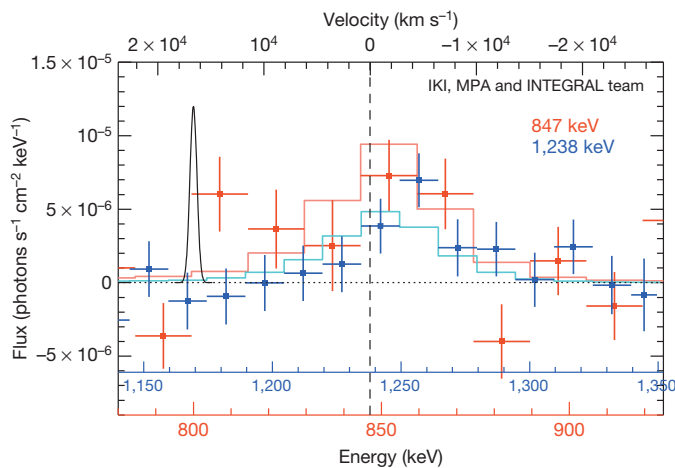


Figure 4 | Broadening of the 847 and 1,238 keV lines. Red points show the SPI spectrum in the 720–920 keV range. The red histogram shows the line profile in the fiducial model. The blue points show the SPI spectrum of the 1,238 keV line. For comparison, a Gaussian line at 800 keV with a width corresponding to the SPI intrinsic energy resolution is shown with a black line. Both observed lines are clearly broadened. The upper axis shows the velocity needed to shift the line to a given energy. All error bars, 1 s.d.

model, the mass-weighted root mean squared velocity of the ejecta is $\sim \sqrt{12}V_e = 10,000 \pm 3,000 \text{ km s}^{-1}$ (Methods and Extended Data Fig. 5). The confidence contours in the $M_{\text{Ni}}-M_{\text{Ej}}$ plane are elongated such that the highest and lowest allowed values of M_{Ni} respectively correspond to the highest and lowest values of M_{Ej} .

In more realistic models, based on calculations of explosive nucleosynthesis, the parameters are not independent and the distribution of elements over the ejecta can vary strongly. We therefore compared the expected spectra for several representative models of type Ia supernovae at day 75, scaled to the distance of M82, with the data. The list of models and corresponding $\Delta\chi^2$ values with respect to the null hypothesis of no source are given in Extended Data Table 2. Remarkably, the canonical model of type Ia supernovae, W7¹, provides the best description of the SN 2014J spectrum, with $\Delta\chi^2 = 54.4$. The pure-detonation model DETO²⁸ produces too much ^{56}Ni and can be reliably rejected. The sub-Chandrasekhar model HeD6²⁹ instead produces a γ -ray flux that is too low and therefore can also be rejected. Our best-fitting three-parameter model (3PAR), and delayed-detonation models DD4³⁰ and DDT1p (E.B. *et al.*, manuscript in preparation), designed to approximately match SN 2014J properties in the visual band, offers a gain in $\Delta\chi^2$ nearly as good as W7. From the standpoint of purely statistical errors, W7 performs significantly better than other models. However, given the inevitable assumptions and simplification in each model, and allowing for possible systematic uncertainties, this group of models cannot be rejected.

Overall, the good agreement with the canonical models shows that in γ -rays SN 2014J looks like a prototypical type Ia supernova, even though strong and complicated extinction in the optical band makes the overall analysis challenging.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 May; accepted 3 July 2014.

1. Nomoto, K., Thielemann, F.-K. & Yokoi, K. Accreting white dwarf models of Type I supernovae. III - Carbon deflagration supernovae. *Astrophys. J.* **286**, 644–658 (1984).
2. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24**, 205–253 (1986).
3. Hillebrandt, W. & Niemeyer, J. C. Type Ia supernova explosion models. *Annu. Rev. Astron. Astrophys.* **38**, 191–230 (2000).
4. Whelan, J. & Iben, I. Jr. Binaries and supernovae of type I. *Astrophys. J.* **186**, 1007–1014 (1973).

5. Iben, I. Jr & Tutukov, A. V. Supernovae of type I as end products of the evolution of binaries with components of moderate initial mass (M not greater than about 9 solar masses). *Astrophys. J. Suppl. Ser.* **54**, 335–372 (1984).
6. Gilfanov, M. & Bogdan, A. An upper limit on the contribution of accreting white dwarfs to the type Ia supernova rate. *Nature* **463**, 924–925 (2010).
7. Röpke, F. K. *et al.* Constraining type Ia supernova models: SN 2011fe as a test case. *Astrophys. J.* **750**, L19 (2012).
8. Malone, C. M. *et al.* The deflagration stage of Chandrasekhar mass models for type Ia supernovae. I. Early evolution. *Astrophys. J.* **782**, 11 (2014).
9. Moll, R., Raskin, C., Kasen, D. & Woosley, S. E. Type Ia supernovae from merging white dwarfs. I. Prompt detonations. *Astrophys. J.* **785**, 105 (2014).
10. The, L.-S. & Burrows, A. Expectations for the hard X-ray continuum and gamma-ray line fluxes from the type Ia supernova SN 2014J in M82. *Astrophys. J.* **786**, 141 (2014).
11. Sunyaev, R. *et al.* Discovery of hard X-ray emission from supernova 1987A. *Nature* **330**, 227–229 (1987).
12. Dotani, T., Hayashida, K., Inoue, H., Itoh, M. & Koyama, K. Discovery of an unusual hard X-ray source in the region of supernova 1987A. *Nature* **330**, 230–231 (1987).
13. Matz, S. M., Share, G. H., Leising, M. D., Chupp, E. L. & Vestrand, W. T. Gamma-ray line emission from SN1987A. *Nature* **331**, 416–418 (1988).
14. Teegarden, B. J., Barthelmy, S. D., Gehrels, N., Tueller, J. & Leventhal, M. Resolution of the 1,238-keV γ -ray line from supernova 1987A. *Nature* **339**, 122–123 (1989).
15. Isern, J. *et al.* Observation of SN2011fe with INTEGRAL. I. Pre-maximum phase. *Astron. Astrophys.* **552**, A97 (2013).
16. Fossey, S., Cooke, B., Pollack, G., Wilde, M. & Wright, T. Supernova 2014J in M82 = Psn J09554214+6940260. *CBET* **3792** (2014).
17. Zheng, W. & Filippenko, A. V. Prediscovery KAIT/LOSS detections of SN 2014J. *Astron. Telegr.* **5822** (2014).
18. Karachentsev, I. D. & Kashibadze, O. G. Masses of the local group and of the M81 group estimated from distortions in the local velocity field. *Astrophysics* **49**, 3–18 (2006).
19. Winkler, C. *et al.* The INTEGRAL mission. *Astron. Astrophys.* **411**, L1–L6 (2003).
20. Sazonov, S. Y., Lutovinov, A. A. & Krivonos, R. A. Cutoff in the hard X-ray spectra of the ultraluminous X-ray sources HolX-1 and M82 X-1. *Astron. Lett.* **40**, 65–74 (2014).
21. Arnett, W. D. Type I supernovae. I - Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).
22. Margutti, R. *et al.* No X-rays from the very nearby Type Ia SN2014J: constraints on its environment. *Astrophys. J.* **790**, 52 (2014).
23. Amanullah, R. *et al.* The peculiar extinction law of SN2014J measured with The Hubble Space Telescope. *Astrophys. J.* **788**, L21 (2014).
24. Nadyozhin, D. K. The properties of Ni to CO to Fe decay. *Astrophys. J. Suppl. Ser.* **92**, 527–531 (1994).
25. Dwarkadas, V. V. & Chevalier, R. A. Interaction of type Ia supernovae with their surroundings. *Astrophys. J.* **497**, 807–823 (1998).
26. Woosley, S. E., Kasen, D., Blinnikov, S. & Sorokina, E. Type Ia supernova light curves. *Astrophys. J.* **662**, 487–503 (2007).
27. Sunyaev, R. A. *et al.* Hard X-radiation from supernova 1987A: Roentgen Observatory observations from 1987 to 1989. *Sov. Astron. Lett.* **16**, 171–176 (1990).
28. Badenes, C., Bravo, E., Borkowski, K. J. & Domínguez, I. Thermal X-ray emission from shocked ejecta in type Ia supernova remnants: prospects for explosion mechanism identification. *Astrophys. J.* **593**, 358–369 (2003).
29. Hoefflich, P. & Khokhlov, A. Explosion models for type Ia supernovae: a comparison with observed light curves, distances, H_0 , and q_0 . *Astrophys. J.* **457**, 500–528 (1996).
30. Woosley, S. E. & Weaver, T. A. in *Supernovae* (eds Audouze, J., Bludman, S., Mochkovitch, R. & Zinn-Justin, J.) 63–154 (Elsevier, 1991).

Acknowledgements This work was based on observations with INTEGRAL, an ESA project with instruments and a science data centre funded by ESA member states (especially the principal investigator countries: Denmark, France, Germany, Italy, Switzerland and Spain) and with the participation of Russia and the United States. We are grateful to the ESA INTEGRAL team and E. Kuulkers for their prompt reaction to the SN 2014J event. E.C., R.S. and S.G. wish to thank the Russian INTEGRAL advisory committee for allocating an additional 10^6 s of time from a regular programme to SN 2014J observations. R.S., S.G. and S.S. are partly supported by grant no. 14-22-00271 from the Russian Scientific Foundation. J.I. is supported by MINECO-FEDER and Generalitat de Catalunya grants. The SPI project has been completed under the responsibility and leadership of CNES, France. ISGRI has been realized by CEA with the support of CNES. We thank P. Höflich, K. Nomoto and S. Woosley for making available their supernova explosion models HED6, W7 and DD4.

Author Contributions E.C.: reduction and modelling of the INTEGRAL observations, simulations of the emerging γ -ray emission, interpretation, manuscript preparation; R.S.: principal investigator of one of the observations, observation planning, interpretation and manuscript preparation; J.I.: principal investigator of one of the observation proposals, observation planning, modelling of observations, interpretation and manuscript preparation; J.K.: reduction and analysis of SPI data, manuscript review; P.J.: reduction and spectral analysis of SPI observations, manuscript review; F.L.: reduction of the IBIS/ISGRI observations, manuscript preparation; N.C.: determination of ^{56}Ni mass from the optical data, manuscript review; S.G.: reduction of the IBIS/ISGRI observations, manuscript review; E.B.: theoretical modelling, manuscript review; S.S.: principal investigator of the M82 observations, manuscript review; M.R.: reduction of the IBIS/ISGRI observations, manuscript review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.C. (churazov@mpa-garching.mpg.de).

METHODS

Observations. INTEGRAL is an ESA scientific mission dedicated to fine spectroscopy and imaging of celestial γ -ray sources in the energy range from 15 keV to 10 MeV. The INTEGRAL data used here were accumulated during revolutions 1391–1407 (ref. 31), corresponding to the period ~ 50 –100 days after the explosion (proposals: 1170002, PI: R.S.; 1140011, PI: J.I.; 1170001, public). Observations were performed in a standard 5×5 pattern around the nominal target location: one source on-axis pointing, 24 ‘off-source pointings’, with a 2.17° step. During off-source pointings, the source remains well within the INTEGRAL field of view. Periods of very high and variable background due to solar flares were omitted from the analysis (Extended Data Fig. 1). The total exposure of the clean data set is ~ 2.6 Ms.

SPI data analysis. SPI³² is a coded-mask germanium spectrometer on board INTEGRAL. The instrument consists of 19 individual Ge detectors, and has a field of view of $\sim 30^\circ$ (at zero response), an effective area of $\sim 70 \text{ cm}^2$ at 0.5 MeV and an energy resolution of $\sim 2 \text{ keV}$. The effective angular resolution of SPI is $\sim 2^\circ$. During SN 2014J observations, 15 out of 19 detectors were operating, resulting in slightly reduced sensitivity and imaging capabilities compared with the initial configuration. The data analysis follows the scheme implemented for the analysis of the Galactic Centre positron annihilation emission^{33,34}. For each detector, a linear relation between the energy and the channel number was assumed and calibrated (separately for each orbit), using the observed energies of background lines (Extended Data Fig. 2) at $\sim 198, 438, 584, 882, 1,764, 1,779, 2,223$ and $2,754 \text{ keV}$. These bright lines are identified³⁵ with known nuclear transitions in Ge and other elements, activated by heavy particles, bombarding the SPI detector. For our analysis, we used a combination of single and pulse-shape-discriminator (PSD) events³², and treated them in the same way.

The count rate from the supernova at energy E , $S(E)$, and the background rates in individual detectors, $B_i(E, t)$, were derived from a simple model of the observed rates, $D_i(E, t)$, in individual SPI detectors, where i is the detector number and t is the time of observation with a typical exposure of 2,000 s: $D_i(E, t) \approx S(E) \times R_i(E, t) + B_i(E, t) + C_i(E)$. Here $R_i(E, t)$ is the effective area for the i th detector, as seen from the source position in a given observation, and $C_i(E)$ does not depend on time. The background rate is assumed to be linearly proportional to the Ge detectors’ saturated event rate, $G_{\text{sat}}(t)$, above 8 MeV, averaged over all detectors; that is, $B_i(E, t) = \beta_i(E)G_{\text{sat}}(t)$. The coefficients $S(E)$, $\beta_i(E)$ and $C_i(E)$ are free parameters of the model and are obtained by minimizing χ^2 for the entire data set. Even though the number of counts in individual exposures is low, it is still possible to use a plain χ^2 approach as long as the errors are estimated using the mean count rate and the total number of counts in the entire data set is large³⁶. The linear nature of the model allows for straightforward estimation of statistical errors.

As an example, we consider SPI data in the 1,200–1,300 keV range, containing the 1,238 keV line of ^{56}Co . For the entire data set there are in total 14,340 flux measurements by individual SPI detectors with a typical exposure time of $\sim 2,000$ s. Fitting this data set with a constant background model (that is, $D_i(E, t) \approx C_i(E)$) yields the reduced $\chi^2_r = 1.31$ per degree of freedom. Adding a $B_i(E, t) = \beta_i(E)G_{\text{sat}}(t)$ term to the model reduces χ^2_r to 0.998, consistent with expectations for photon counting noise. Further adding one more parameter—the flux of a source at the position of SN 2014J—lowers χ^2_r by $\Delta\chi^2 = 16.4$, corresponding to a ~ 4 s.d. detection. For comparison, for the 600–800 keV band, which does not contain strong lines, the improvement in χ^2 by adding the source flux to the model is only 0.60, giving no significant evidence for SN 2014J emission.

Despite its proximity, SN 2014J is still an extremely faint source in γ -rays. Although precise measurements of line fluxes are challenging, a combination of spectral and imaging information makes our results very robust. To further test the possible influence of variable background on our results, we repeated the calculation of the SN 2014J spectrum while dropping the $B_i(E, t)$ term from the model. The line fluxes changed by $\sim 3\%$ compared with our more elaborate baseline background model. However, the wings of the lines, where the background is $\sim 10^4$ times higher than the source, can still suffer from the residual background features. The uncertainty in the distance $D = 3.53 \pm 0.26 \text{ Mpc}$ (ref. 18) causes an additional $\sim 15\%$ uncertainty in the flux and, therefore, in the estimated mass of radioactive Ni and Co. **ISGRI/IBIS data analysis.** The primary imaging instrument on board INTEGRAL is IBIS³⁷, which is a coded-mask aperture telescope with the CdTe-based detector ISGRI³⁸. It has higher sensitivity to continuum emission than SPI in the 20–300 keV range and has a spatial resolution of $\sim 12'$. We note here that neither ISGRI nor SPI can distinguish the emission of SN 2014J from the emission of any other source in M82. ISGRI, however, can easily differentiate between M82 and M81, which are separated by $\sim 30'$. The energy resolution of ISGRI is $\sim 10\%$ at 100 keV. The ISGRI energy calibration uses the procedure implemented in OSA 10.0 (ref. 39). The images in broad energy bands were reconstructed using a standard mask–detector cross-correlation procedure, tuned to produce zero signal on the sky if the count rate across the detector matches the pattern expected from pure background, which was derived from the same data set by stacking detector images. The noise in the

resulting images is fully consistent with the expected level, determined by photon counting statistics. The fluxes in broad bands were calibrated using Crab Nebula observations with INTEGRAL shortly before the observations of SN 2014J discussed here.

Fitting 847 and 1,238 keV line parameters. The flux, energy centroid and broadening of the lines were evaluated by fitting a Gaussian to portions of the SPI spectrum in the 800–900 and 1,100–1,350 keV bands. The best-fit parameters are given in Extended Data Table 1 along with 1 s.d. uncertainties. Although for plotting purposes we used heavily binned spectra (Figs 1 and 4), the fitting is done for the unbinned SPI spectrum in 0.5 keV-wide channels.

Because the decay time of ^{56}Co ($\tau = 111.4$ days) and branching ratios (1 and 0.66 for the 847 and 1,238 keV lines, respectively) are known²⁴, it is straightforward to convert line fluxes into the mass of ^{56}Co visible to INTEGRAL at the time of observation: $M_{\text{Co}} = (0.16 \pm 0.05)M_\odot$ and $(0.27 \pm 0.07)M_\odot$ for the 847 and 1,238 keV lines, respectively. These values can be considered model-independent lower limits on the amount of ^{56}Co at day 75 since the explosion. The fractions of line photons escaping the ejecta without interactions were estimated from our fiducial model as 0.60 (847 keV) and 0.64 (1,238 keV). These values were used to correct the observed fluxes to derive estimates of the total ^{56}Co mass at day 75, $M_{\text{Co}} = (0.26 \pm 0.08)M_\odot$ and $(0.42 \pm 0.11)M_\odot$, on the basis of the 847 and 1,238 keV line fluxes, respectively. The derived masses are consistent within the uncertainties with a mean value of $M_{\text{Co}} = (0.34 \pm 0.07)M_\odot$. Finally, a correction factor of 1/0.55 has been applied to convert the mass of ^{56}Co at day 75 to the initial mass of ^{56}Ni : $M_{\text{Ni}} = (0.6 \pm 0.13)M_\odot$. This factor accounts for time evolution of the Co mass in the decay chain from Ni to Fe.

In fully transparent ejecta, the centroid of emerging γ -ray lines should be unshifted (at least to first order in V_e/c , where c is the speed of light). Opacity suppresses γ -rays coming from the receding part of the ejecta, leading to a blueshift of the visible line. Blueshift is indeed observed for both lines (Fig. 4).

The corresponding mean velocity is $V_{\text{shift}} = -3,100 \pm 1,100 \text{ km s}^{-1}$. This value is slightly higher than the expected shift of $-1,280 \text{ km s}^{-1}$, estimated from the fiducial model for γ -ray photons escaping the ejecta without interactions.

The expected line broadening (root mean squared line-of-sight velocity), σ_γ , for transparent ejecta is directly related to the characteristic expansion velocity.

In the model with an exponential density profile, $\sigma_\gamma = 2V_e$. Indeed, in the model we found for directly escaping photons, $\sigma_\gamma = 5,860 \text{ km s}^{-1} \approx 2.1V_e$. The Gaussian fit to the observed lines yields a slightly lower value, $\sigma_\gamma = 4,100 \pm 960 \text{ km s}^{-1}$. It is possible that the exact values of the shift and broadening are affected by remaining uncertainties in the background model. However, the presence of shift and broadening at the level of few 10^3 km s^{-1} is very robust.

Simplified model. The emergent γ -ray spectrum from a type Ia supernova is determined by interactions of γ -ray quanta with the expanding ejecta and can be used as a proxy for the most basic properties of the supernova¹⁵. Although detailed modelling of the properties of SN 2014J is beyond the scope of this Letter, we use a simple model to qualitatively compare our results with expectations. Our basic observables are the line fluxes, primarily determined by M_{Ni} ; line broadening, set by the typical energy release per unit mass; and the flux below 511 keV, which is sensitive to the Thompson depth of the ejecta. Accordingly, we build a spherically symmetric model of homologously expanding ejecta with mass M_{Ej} , ignoring the possible anisotropy of the ^{56}Ni distribution expected in the scenario of white dwarf mergers⁹. The density follows an exponential law^{25,26} $\rho \propto e^{-v/V_e}$, where $V_e = \sqrt{E_K/6M_{\text{Ej}}}$ and is truncated at $v = 10V_e$, and E_K is the kinetic energy of the ejecta. In this model a mass-weighted root-mean-squared velocity of the ejecta is $\sqrt{12}V_e$. All elements, including radioactive Ni and Co, are uniformly mixed through the entire ejecta, which are composed of iron group elements (58%) and Si and S in equal proportions (21% each). This is of course a strong simplification, certainly violated in the outer regions and in the core, but it allows us to specify the model completely with only three parameters: M_{Ni} , M_{Ej} and V_e . The presence of the radioactive elements in the outer layers is crucial for the early phase of the supernova evolution, but becomes less important for days 50–100, which are the days of interest here. As a fiducial example, we use $M_{\text{Ni}} = 0.7M_\odot$, $M_{\text{Ej}} = 1.38M_\odot$ and $V_e = 2,800 \text{ km s}^{-1}$, corresponding to $E_K = 1.3 \times 10^{51} \text{ erg}$.

A Monte Carlo radiative transfer code is used to calculate the emergent spectrum, which includes full treatment of Compton scattering (coherent and incoherent) and photoabsorption. Pair production by γ -ray photons is neglected. The positrons produced by β^+ decay of ^{56}Co (19% of all decays) annihilate in place via positronium formation. Both two-photon annihilation into the 511 keV line and the ortho-positronium continuum are included. Our reference model was calculated for day 75 since the explosion. A time delay due to the finite propagation time of the photons is neglected (it amounts to few days from the radius where the bulk of the mass is located).

Examples of the model spectra for days 50, 75 and 100 are shown in Extended Data Fig. 3. To a first approximation (justified by the low signal-to-noise ratio of

the observed spectrum), the model spectrum for day 75 can be used for comparison with the observed spectrum, accumulated by INTEGRAL over days 50–100. The contributions of various components to the total spectrum are shown in Extended Data Fig. 4. The most prominent are the lines at 847 and 1,238 keV. These lines are broadened by the expansion velocity of the ejecta and escape without any interactions (the role of coherent scattering is negligible at these energies). The line shape is also modified by opacity effects, because γ -ray photons produced in the nearest (approaching) side of the ejecta have a higher chance of reaching an observer. This effect causes the line shapes to be skewed towards the blue side. The scattered continuum associated with most prominent lines extends down to energies ~ 100 keV. At lower energies, photoabsorption becomes dominant and the flux drops. In addition, three-photon annihilation, associated with the decay of ortho-positronium, produces a significant contribution to the continuum flux. Unlike the scattered continuum, this component does not scale with the Thomson depth of the ejecta and becomes progressively more important for late evolution of the emergent spectrum.

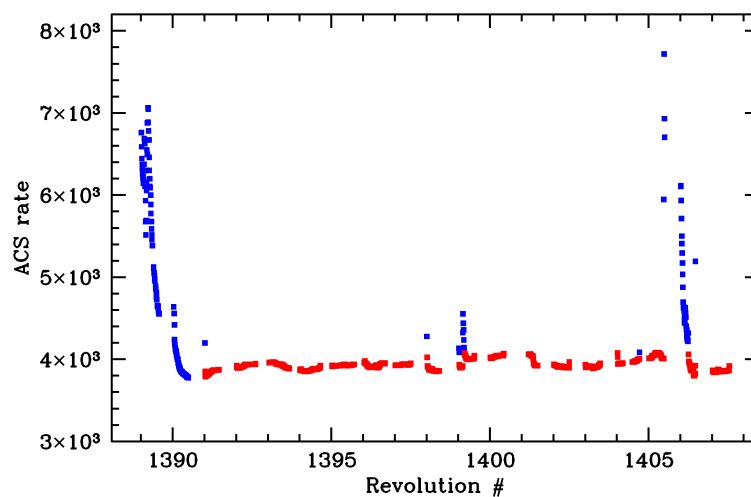
With our simple three-parameter model it is possible to run a grid of models to evaluate a plausible range of basic parameters. This was done by varying M_{Ni} , M_{Ej} and V_{e} , comparing the model with the spectrum and calculating χ^2 . It is convenient to express the success of the model by the reduction of $\Delta\chi^2$ relative to the null hypothesis of no source. The null hypothesis gives $\chi^2 = 1,948.7$ for 1,905 degrees of freedom for the combined SPI+ISGRI spectrum. The best-fitting three-parameter model has $\Delta\chi^2 = 50.5$, corresponding to a 7.1 s.d. detection with $M_{\text{Ni}} = 0.56 M_{\odot}$, $M_{\text{Ej}} = 1.2 M_{\odot}$ and $V_{\text{e}} = 3,000 \text{ km s}^{-1}$, corresponding to $E_{\text{K}} = 1.3 \times 10^{51} \text{ erg}$. The same model allows for calculation of 1 s.d. confidence intervals (for a single parameter of interest) by identifying parameter space, which has $\Delta\chi^2$ smaller by 1 than the best-fitting model. Corresponding confidence intervals are shown in Extended Data Fig. 5.

Specific explosion models. Apart from our simplified three-parameter model, we compare the spectrum (SPI+ISGRI) with the expected spectra calculated (E.B. *et al.*, manuscript in preparation) for several detailed type Ia supernova explosion models. For each model, we calculate the reduction of $\Delta\chi^2$ relative to the null hypothesis of no source. The resulting values and basic characteristics of the models are given in Extended Data Table 2. The set of models includes the classic W7 (ref. 1) and DD4 (ref. 30) models, the pure-detonation model, DETO²⁸, the sub-Chandrasekhar model, HED6 (ref. 29), our three-parameter model 3PAR with fiducial and best-fitting parameters, and several variants of the delayed-detonation model DDT1p (E.B. *et al.*, manuscript in preparation). DDT1p4 was built to match SN 2014J in the visible band. DDT1p1 is a slightly less energetic version of DDT1p4. In the DDT1p4halo model, the object is surrounded by a $0.2 M_{\odot}$ ‘halo’, which can be envisaged in a slow merger scenario.

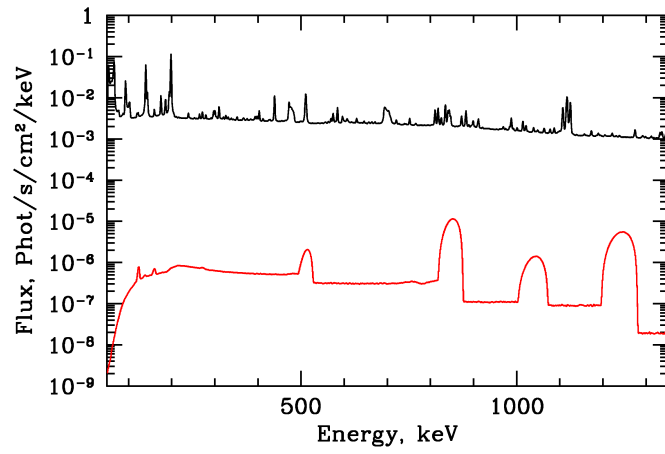
Although the single-white-dwarf models discussed above provide consistent descriptions of the INTEGRAL data, these observations by themselves do not immediately exclude double-white-dwarf merger scenarios, provided that similar amounts of ^{56}Ni are synthesized. In this regard, we note that no direct evidence that the progenitor of SN 2014J was a single accreting white dwarf has been found so far, placing tight limits on the most popular accretion scenarios^{22,40–42}.

Optical bolometric luminosity during INTEGRAL observations. Using recent BVRIJHK photometry⁴³, we estimated the supernova optical bolometric luminosity on day 75 after explosion (median of INTEGRAL observations). For $A_V = 1.85 \text{ mag}$ and $R_V = 1.6$, we found the bolometric luminosity of $\sim 1.1 \times 10^{42} \text{ erg s}^{-1}$, which agrees well with the estimates of deposited power in our fiducial model.

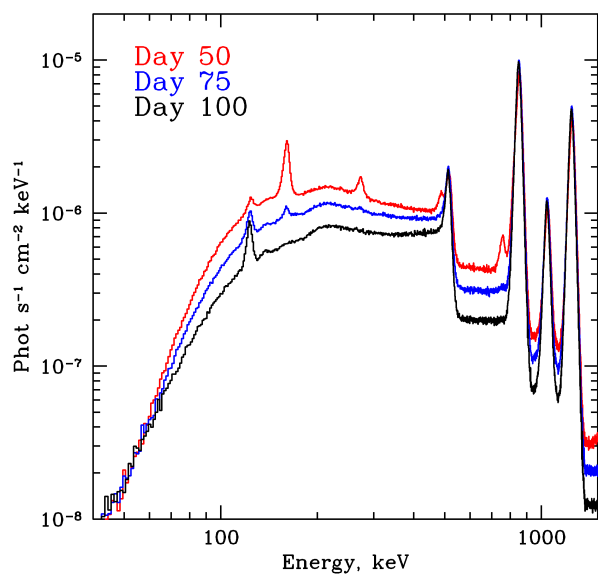
31. Kuulkers, E. INTEGRAL target of opportunity observations of the type Ia SN2014J in M82. *Astron. Telegr.* **5835** (2014).
32. Vedrenne, G. *et al.* SPI: the spectrometer aboard INTEGRAL. *Astron. Astrophys.* **411**, L63–L70 (2003).
33. Churazov, E., Sunyaev, R., Sazonov, S., Revnivtsev, M. & Varshalovich, D. Positron annihilation spectrum from the Galactic Centre region observed by SPI/INTEGRAL. *Mon. Not. R. Astron. Soc.* **357**, 1377–1386 (2005).
34. Churazov, E., Sazonov, S., Tsygankov, S., Sunyaev, R. & Varshalovich, D. Positron annihilation spectrum from the Galactic Centre region observed by SPI/INTEGRAL revisited: annihilation in a cooling ISM? *Mon. Not. R. Astron. Soc.* **411**, 1727–1743 (2011).
35. Weidenspointner, G. *et al.* First identification and modelling of SPI background lines. *Astron. Astrophys.* **411**, L113–L116 (2003).
36. Churazov, E., Gilfanov, M., Forman, W. & Jones, C. Mapping the gas temperature distribution in extended X-ray sources and spectral analysis in the case of low statistics: application to ASCA observations of clusters of galaxies. *Astrophys. J.* **471**, 673–682 (1996).
37. Ubertini, P. *et al.* IBIS: the imager on-board INTEGRAL. *Astron. Astrophys.* **411**, L131–L139 (2003).
38. Lebrun, F. *et al.* ISGRI: the INTEGRAL Soft Gamma-Ray Imager. *Astron. Astrophys.* **411**, L141–L148 (2003).
39. INTEGRAL Science Data Centre. *INTEGRAL* <http://isdc.unige.ch/integral> (2014).
40. Nielsen, M. T. B., Gilfanov, M., Bogdán, Á., Woods, T. E. & Nelemans, G. Upper limits on the luminosity of the progenitor of type Ia supernova SN 2014J. *Mon. Not. R. Astron. Soc.* **442**, 3400–3406 (2014).
41. Kelly, P. L. *et al.* Constraints on the progenitor system of the type Ia supernova 2014J from pre-explosion Hubble Space Telescope imaging. *Astrophys. J.* **790**, 3 (2014).
42. Perez-Torres, M. A. *et al.* Constraints on the progenitor system and the environs of SN 2014J from deep radio observations. Preprint at <http://arxiv.org/abs/1405.4702> (2014).
43. Foley, R. J. *et al.* Extensive HST ultraviolet spectra and multi-wavelength observations of SN 2014J in M82 indicate reddening and circumstellar scattering by typical dust. Preprint at <http://arxiv.org/abs/1405.3677> (2014).



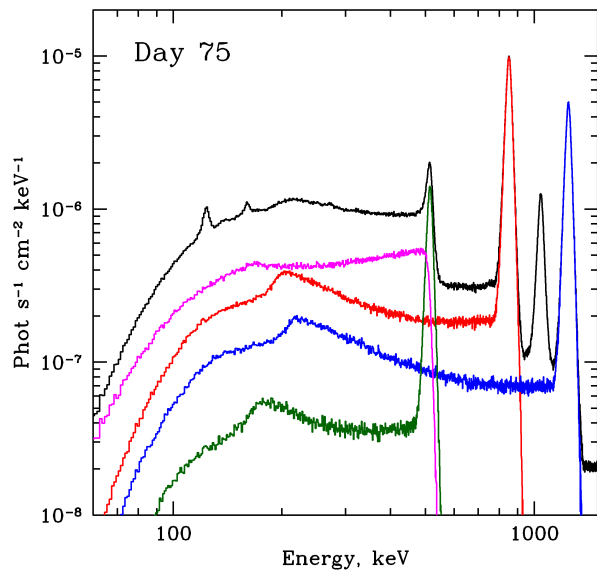
Extended Data Figure 1 | Variations in the particle background during INTEGRAL observations. Anti-coincidence-system count rate is shown as a function of time, expressed through the revolution number. One revolution lasts about 3 days. Periods of very high and variable background (shown in blue) due to solar flares were omitted from the analysis. Periods of quiescent background (red) were used to derive the spectrum of SN 2014J.



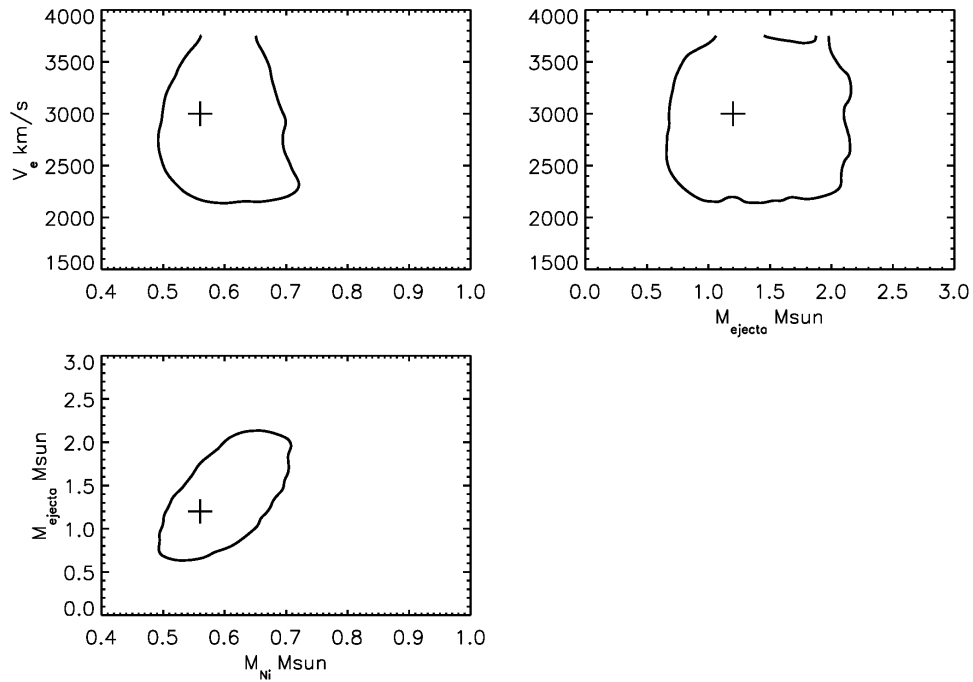
Extended Data Figure 2 | Comparison of the SPI background spectrum and the expected type Ia supernova emission. Typical quiescent background (black) and supernova model (red, convolved with SPI energy resolution) spectra.



Extended Data Figure 3 | Predicted spectra for days 50, 75 and 100 after explosion. The 3PAR model spectrum calculated for day 75 is used for comparison with the INTEGRAL data obtained between day 50 and 100 since the explosion. Weak lines below 200 keV correspond to ^{56}Ni (day 50) and ^{57}Co (day 100).



Extended Data Figure 4 | Contributions of various components to the model spectrum. The lines are formed by γ -rays escaping the ejecta without interactions. The low-energy tail of each line is due to Compton down-scattering of the photons because of the recoil effect. The 'humps' in the tails correspond to the scattering by 180° . The magenta line shows the contribution of the ortho-positronium annihilation. Annihilation of para-positronium contributes to 511 keV line.



Extended Data Figure 5 | Confidence contours for our three-parameter model. The cross shows the best-fit values. Contours are plotted at $\Delta\chi^2 = 1$ with respect to the best-fit value and characterize the 1 s.d. confidence interval for a single parameter of interest.

Extended Data Table 1 | Parameters of the observed gamma-ray lines of ⁵⁶Co

Parameter	847 keV line	1238 keV line
Flux, 10 ⁻⁴ phot cm ⁻² s ⁻¹	2.34±0.74	2.78±0.74
Luminosity, 10 ⁴¹ erg s ⁻¹	4.7	8.1
V _{shift} (l.o.s. velocity), km s ⁻¹	-1900±1600	-4300±1600
Line width (l.o.s. velocity rms), σ _v , km s ⁻¹	3600±1300	4700±1400

Each line is independently approximated with a Gaussian. Errors, 1 s.d.

Extended Data Table 2 | Comparison of typical type Ia supernova explosion models with the data

Model	$M_{\text{Ni}}, M_{\text{Sun}}$	$M_{\text{Ej}}, M_{\text{Sun}}$	$E_{\text{K}}, 10^{51} \text{ erg}$	$\Delta\chi^2$
W7 ¹	0.59	1.38	1.24	54.4
DDT1p1	0.54	1.36	1.29	52.5
DD4 ³⁰	0.61	1.39	1.24	52.0
DDT1p4	0.66	1.36	1.35	51.9
3PAR, best-fitting	0.56	1.20	1.3	50.5
3PAR, fiducial	0.70	1.38	1.3	49.3
DDT1p4halo	0.62	1.55	1.3	49.1
HED6 ²⁹	0.26	0.77	0.72	38.2
DETO ²⁸	1.16	1.38	1.44	12.1

The rightmost column shows the improvement in χ^2 , relative to the null hypothesis of no source, by assuming a type Ia supernova spectral model (no free parameters). The null hypothesis gives $\chi^2 = 1,948.7$ for 1,905 degrees of freedom.

Quantum imaging with undetected photons

Gabriela Barreto Lemos^{1,2}, Victoria Borish^{1,3}, Garrett D. Cole^{2,3}, Sven Ramelow^{1,3,†}, Radek Lapkiewicz^{1,3} & Anton Zeilinger^{1,2,3}

Information is central to quantum mechanics. In particular, quantum interference occurs only if there exists no information to distinguish between the superposed states. The mere possibility of obtaining information that could distinguish between overlapping states inhibits quantum interference^{1,2}. Here we introduce and experimentally demonstrate a quantum imaging concept based on induced coherence without induced emission^{3,4}. Our experiment uses two separate down-conversion nonlinear crystals (numbered NL1 and NL2), each illuminated by the same pump laser, creating one pair of photons (denoted idler and signal). If the photon pair is created in NL1, one photon (the idler) passes through the object to be imaged and is overlapped with the idler amplitude created in NL2, its source thus being undefined. Interference of the signal amplitudes coming from the two crystals then reveals the image of the object. The photons that pass through the imaged object (idler photons from NL1) are never detected, while we obtain images exclusively with the signal photons (from NL1 and NL2), which do not interact with the object. Our experiment is fundamentally different from previous quantum imaging techniques, such as interaction-free imaging⁵ or ghost imaging^{6–9}, because now the photons used to illuminate the object do not have to be detected at all and no coincidence detection is necessary. This enables the probe wavelength to be chosen in a range for which suitable detectors are not available. To illustrate this, we show images of objects that are either opaque or invisible to the detected photons. Our experiment is a prototype in quantum information—knowledge can be extracted by, and about, a photon that is never detected.

The conceptual arrangement of our imaging technique, based on a quantum interference experiment^{3,4} by Zou, Wang and Mandel, is illustrated in Fig. 1. A pump beam (green) divided by a 50:50 beam splitter (BS1) coherently illuminates two identical nonlinear crystals, NL1 and NL2, where pairs of collinear photons called signal (yellow) and idler (red) can be created ($|c\rangle_s|d\rangle_i$ in NL1 and $|e\rangle_s|f\rangle_i$ in NL2). The idler amplitude created in NL1 reflects at the dichroic mirror D1 into spatial mode d , and signal amplitude passes into spatial mode c . The idler passes through the object O of real transmittance coefficient T and phase shift γ : $|c\rangle_s|d\rangle_i \rightarrow Te^{i\gamma}|c\rangle_s|d\rangle_i + \sqrt{1-T^2}|c\rangle_s|w\rangle_i$, where for simplicity we lump all lost idler amplitude into a single state $|w\rangle_i$ (here subscripts s and i represent signal and idler). By reflection at dichroic mirror D2, the idler from NL1 aligns perfectly with idler amplitude produced at NL2, $|d\rangle_i \rightarrow |f\rangle_i$. The state at the grey dotted line is thus

$$\frac{1}{\sqrt{2}} \left[(Te^{i\gamma}|c\rangle_s + |e\rangle_s)|f\rangle_i + \sqrt{1-T^2}|c\rangle_s|w\rangle_i \right] \quad (1)$$

The idler is now reflected at the dichroic mirror D3 and discarded. The signal states $|c\rangle_s$ and $|e\rangle_s$ are combined at the 50:50 beam splitter BS2. The detection probabilities at the outputs $|g\rangle_s$ and $|h\rangle_s$, obtained by ignoring (tracing out) the idler modes, are

$$P_{g/h} = \frac{1}{2} [1 \pm T \cos \gamma] \quad (2)$$

Thus, fringes with visibility T can be seen at either output, even though the signals combined at BS2 have different sources^{4,10}. These fringes appear in

the signal single photon counts; the idlers are not detected. No coincidence detection is required.

The peculiar feature of this interferometer is that no detected photon has taken path d . Yet, in our experiment, it is precisely here where we put the object to be imaged. The key to this experiment is how the signal-source information carried by the undetected idler photon depends on T . For, if $T = 0$, an idler detected after D3, coincident with a signal count at $|g\rangle_s$ or $|h\rangle_s$, would imply the signal source was NL2. Detection of a signal photon without a coincident idler would imply the signal source was NL1. This which-source information destroys interference because it makes the quantum states overlapping at BS2 distinguishable. If $T = 1$, the idler photon carries no which-source information. The signal states overlapped at each output of BS2 are then indistinguishable; thus the interference term in equation (2) appears. The above arguments are valid even though the idler photons are not detected, for it is only the possibility of obtaining which-source information that matters in this experiment.

Our experiment has a connection to interaction-free measurements^{11,12}. Note that $P_h = 0$ if no object is placed in the set-up ($T = 1$ and $\gamma = 0$). Now insert an opaque object ($T = 0$) so that $P_h > 0$, and monitor the idler reflection from D3. Coincident counts in $|h\rangle_s$ and the idler detector reveal that the object is present even though no photon interacted with the object. With our set-up it is thus possible to realize non-degenerate interaction-free imaging.

With O and D2 removed, equation (1) would be an ordinary two-particle entanglement¹³, $|c\rangle_s|d\rangle_i + |e\rangle_s|f\rangle_i$. With them in, $|d\rangle_i \rightarrow Te^{i\gamma}|f\rangle_i$, which creates equation (1). A normal two-particle entanglement has changed into an interesting single-particle superposition, which is especially rich when T and γ are transverse-position dependent.

We expand the conceptual arrangement of Fig. 1 into an imaging system (Fig. 2). We replace the photon counters with cameras sensitive to single photons and the uniform object with one bearing features, that is, $T = T(x, y)$ and $\gamma = \gamma(x, y)$ depend on transverse position (x, y) . Our source produces spatially entangled photon pairs^{14,15}. Sharp spatial correlations between signal and idler in the object plane and confocal lens systems¹⁶ (see Methods) guarantee a point-by-point correspondence between the object plane and the detector surface on the camera.

The intensity image (non-constant transmittance) is due to transverse-position-dependent which-source information carried by the undetected idler photons. The phase image is of a different nature: it is due to the fact that the position-dependent phase shift on the idler photons in path d is actually passed to the signal; that is¹⁷, $|c\rangle_s (Te^{i\gamma}|f\rangle_i) + |e\rangle_s|f\rangle_i = (Te^{i\gamma}|c\rangle_s + |e\rangle_s)|f\rangle_i$. Remarkably, the idler beam $|f\rangle_i$ alone does not even carry the phase pattern, and without detection in coincidence it could not be used to obtain the phase image^{18,19}.

We will now show images obtained by detecting 810-nm photons with a camera capable of single-photon sensitivity at this wavelength, when three different objects are illuminated by 1,550-nm photons, to which our camera is blind (see Methods). First, a cardboard cut-out placed into the path D1–D2 is imaged. Next, we show that a position-dependent phase shift produces an image even when the object is opaque (an etched silicon plate) or invisible (etched silica plate) at the detection

¹Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, Vienna A-1090, Austria. ²Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, A-1090 Vienna, Austria. ³Quantum Optics, Quantum Nanophysics, Quantum Information, University of Vienna, Boltzmanngasse 5, Vienna A-1090, Austria. [†]Present address: Cornell University, 159 Clark Hall, 142 Science Drive, Ithaca, New York 14853, USA.

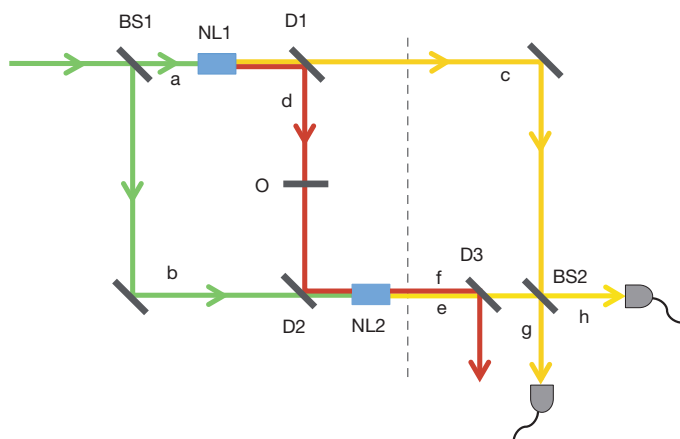


Figure 1 | Schematic of the experiment. Laser light (green) splits at beam splitter BS1 into modes a and b. Beam a pumps nonlinear crystal NL1, where collinear down-conversion may produce a pair of photons of different wavelengths called signal (yellow) and idler (red). After passing through the object O, the idler reflects at dichroic mirror D2 to align with the idler produced in NL2, such that the final emerging idler $|f\rangle$ does not contain any information about which crystal produced the photon pair. Therefore, signals $|c\rangle_s$ and $|e\rangle_s$ combined at beam splitter BS2 interfere. Consequently, signal beams $|g\rangle_s$ and $|h\rangle_s$ reveal idler transmission properties of object O.

wavelength. The images obtained with an electron multiplying charge coupled device (EMCCD) camera show single (non-heralded) counts per pixel ($16\ \mu\text{m} \times 16\ \mu\text{m}$) obtained in an exposure time of 0.5 s with an electron multiplying gain factor of 20. The visibility achieved in the experiment is 77% (see Methods for details).

Figure 3a shows the beamsplitter output when a cardboard cut-out (illustrated in Fig. 3b) is inserted in the path D1–D2. Constructive interference is seen at one output of the beam splitter and destructive interference is observed in the other output. Interference only occurs in the region corresponding to the idler beam transmitted through the shape cut out of the cardboard, as seen in the sum and difference of the complementary images, shown in Fig. 3c and d, respectively. The sum of the two outputs of the beamsplitter gives the featureless intensity profile of the signal beams, demonstrating that the signal beams, while carrying the intensity information, are not absorbed at all by the mask.

In Fig. 4a, we show the image of an etched 500- μm -thick silicon plate; the plate is shown in Fig. 4b (see Methods section for details of the silicon plate and the etching process). Silicon is opaque to illumination at 810 nm,

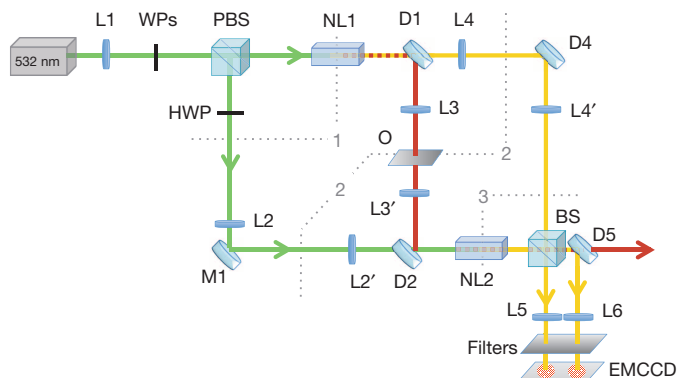


Figure 2 | Experimental set-up. A continuous-wave 532-nm laser (green) illuminates crystals NL1 and NL2. Wave plates (WPs) adjust the relative phase and intensity of the outputs of the polarizing beam splitter (PBS). The dichroic mirror D1 separates down-converted 810-nm (yellow) and 1,550-nm (red) photons. The 1,550-nm photons are transmitted through the object O and sent through NL2 by dichroic mirror D2. Lenses image plane 1 onto plane 3, and plane 2 onto the EMCCD camera. A 50:50 beam splitter (BS) combines the 810-nm beams. Dichroic mirrors D1, D2, D4 and D5 transmit the pump.

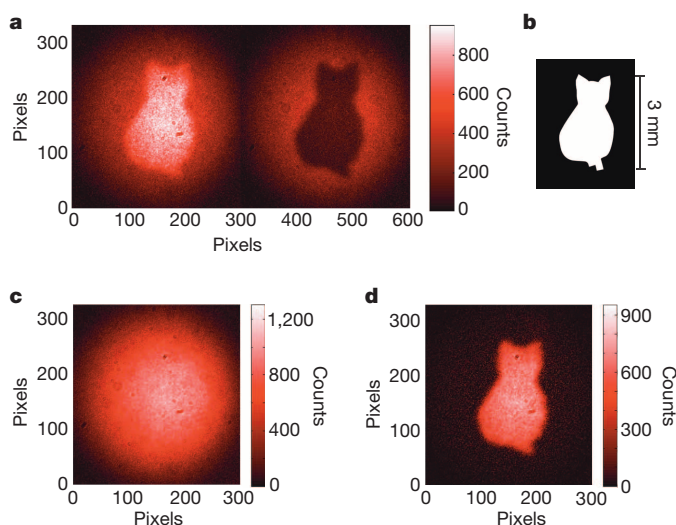


Figure 3 | Intensity imaging. a, Inside the cat, constructive and destructive interference are observed at the outputs of BS when we placed the cardboard cut-out shown in b in the path D1–D2. Outside the cat, idler photons from NL1 are blocked and therefore the signals do not interfere. c, The sum of the outputs gives the intensity profile of the signal beams. d, The subtraction of the outputs leads to an enhancement of the interference contrast, as it highlights the difference between constructive and destructive interference.

thus it is impossible to realize transmission imaging by illuminating the silicon with 810-nm photons. However, silicon is highly transparent at 1,550 nm and when we place the object in path D1–D2, the difference in optical path length for the etched and non-etched regions corresponds to a relative phase shift of π . Even though our camera is blind to 1,550-nm light, the image is seen by detecting 810-nm photons at the output of BS2 (Fig. 4a).

Finally, Fig. 5a shows the image of a fused silica (SiO_2) plate etched with a pattern that is invisible at the detection wavelength (details are given in the Methods section). We take advantage of the flexibility of our source to obtain collinear non-degenerate down-conversion at 820 nm

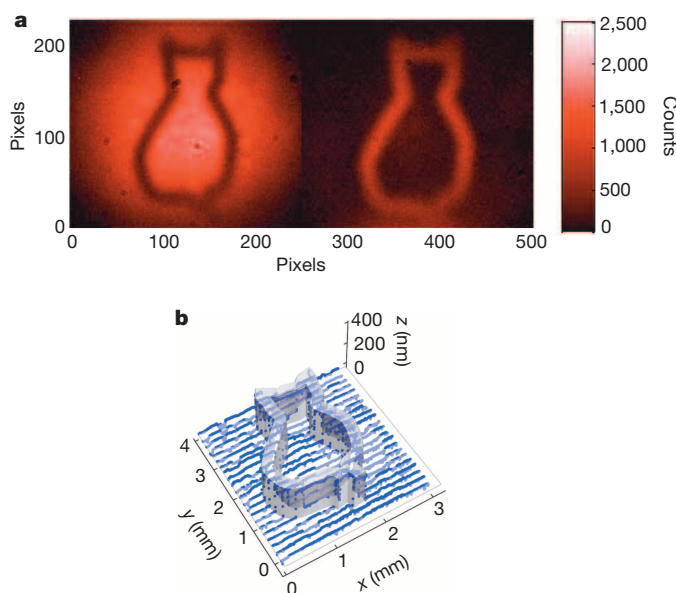


Figure 4 | Phase image of an object opaque to 810-nm light. a, Detection of 810-nm photons at both outputs of BS when a silicon plate (opaque to 810-nm light) with a 3-mm-tall etched cat (b) was introduced in path D1–D2. b, Three-dimensional rendering of the etch design overlaid with stylus profilometer scans (blue points) of the actual etch depth.

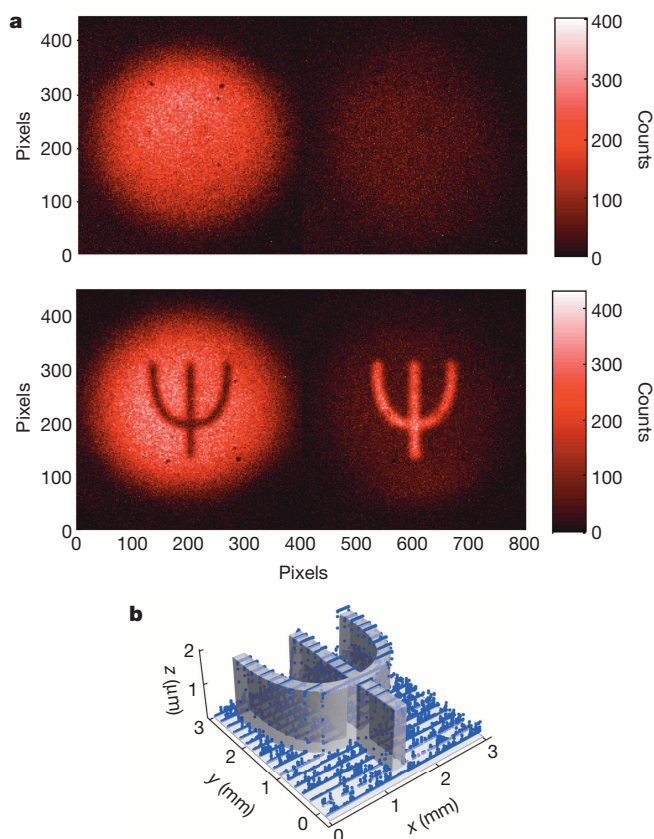


Figure 5 | Phase imaging of a 2π step at 820 nm. **a**, The top picture was taken with the object (shown in **b**) placed in the 820-nm beam between L4 and L4'; in the bottom picture, the object was placed in the 1,515-nm beam in path D1–D2. **b**, Three-dimensional rendering of the design overlaid with stylus profilometer scans (blue dots) of the actual etch depth.

and 1,515 nm (see Methods). The object (Fig. 5b) has an etch depth of 1,803 nm, imparting a relative phase shift of $\sim 2\pi$ for 820-nm light. Thus the object is invisible when placed in the path of the detected photons between L4 and L4' (top of Fig. 5a). This same etch depth gives an $\sim \pi$ phase step for 1,515-nm light, so when this same object is placed in the path D1–D2 of undetected photons, an image seen in the contrast of constructive to destructive interference is retrieved in the 820-nm outputs (bottom of Fig. 5a).

In summary, we have presented a quantum system for intensity and phase imaging where the photons that illuminate the object are not detected and the photons that are detected do not illuminate the object. We image objects that are either opaque or invisible at the detection wavelength (near-infrared) by illuminating three different objects with a wavelength to which our detector is blind. This experiment is fundamentally different to ghost imaging^{6–9} as it relies on single-photon interference and does not require coincidence detection. Furthermore, our technique could be used for non-degenerate interaction-free imaging, with potential applications spanning biological imaging to the inspection of integrated circuits. Our system can realize grey-scale intensity or phase imaging, and it can be modified in order to measure spectral features (spectral imaging)²⁰.

We have demonstrated that our technique does not require the laser or the detector to function at the same wavelength as that of the light probing the object. Additionally, any nonlinear process can be used as a source, and this provides flexibility in the wavelength range for both detection and illumination of the object. In particular, in spontaneous parametric down-conversion (as used here), the only absolute restriction is that the sum of the two photon energies equals that of the pump photons. We have shown

that information can be obtained about an object without detecting the photons that interacted with the object. Knowing the two-photon state, one can obtain information about an object. It has not escaped our attention that, on the other hand, by knowing the object, one could obtain information about the quantum state without detecting it.

METHODS SUMMARY

A detailed schematic of our imaging set-up is shown in Fig. 2. A 532-nm linearly polarized Gaussian pump laser beam focused by lens L1 on plane 1 is divided at a polarizing beam splitter (PBS) and coherently illuminates two identical periodically poled potassium titanyl phosphate (ppKTP) crystals, NL1 and NL2. The PBS plus wave plates (WPs) are used to control the relative amplitudes and phases between the reflected and transmitted pump beams. With an extra half-wave plate (HWP) in the reflected beam, both beams have the same polarization. The 1,550-nm idler amplitude produced at NL1 is reflected by dichroic mirror D1, through which the 810-nm signal and the pump are transmitted. Dichroic mirror D4 transmits 532-nm light and reflects 810-nm light. A long-pass filter (not shown in the figure) placed directly before the object O blocks any residual 532-nm or 810-nm light. The 1,550-nm amplitude from NL1 illuminates the object O and is then overlapped with the pump beam at dichroic mirror D2 that transmits 532-nm light and reflects 1,550-nm light.

Lens pairs L2–L2', L3–L3', and L4–L4' image plane 1 onto plane 3, thereby ensuring that pump, idler and signal, respectively, are identical in these planes, thus contributing to obtain high interference visibility²¹ (see Methods). Lenses L5 and L6 together with L3' and L4' image object plane 2 onto the camera surface.

The 810 ± 1.5 nm photons are detected (without heralding) in both outputs of the BS using an EMCCD camera that exhibits single-photon sensitivity at 810 nm, but has a negligible response at 1,550 nm.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 January; accepted 11 June 2014.

1. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics* Vol. III, Chs 1 and 3 (Addison-Wesley, 1964).
2. Mandel, L. Coherence and indistinguishability. *Opt. Lett.* **16**, 1882–1883 (1991).
3. Zou, X. Y., Wang, L. J. & Mandel, L. Induced coherence and indistinguishability in optical interference. *Phys. Rev. Lett.* **67**, 318–321 (1991).
4. Wang, L. J., Zou, X. Y. & Mandel, L. Induced coherence without induced emission. *Phys. Rev. A* **44**, 4614–4622 (1991).
5. White, A. G., Mitchell, J. R., Nairz, O. & Kwiat, P. G. “Interaction-free” imaging. *Phys. Rev. A* **58**, 605–613 (1998).
6. Abouraddy, A. F., Stone, P. R., Sergienko, A. V., Saleh, B. E. A. & Teich, M. C. Entangled-photon imaging of a pure phase object. *Phys. Rev. Lett.* **93**, 213903 (2004).
7. Gatti, A., Brambilla, E. & Lugiato, L. Quantum imaging. *Prog. Opt.* **51**, 251–348 (2008).
8. Pittman, T. B. et al. Two-photon geometric optics. *Phys. Rev. A* **53**, 2804–2815 (1996).
9. Aspden, R. S., Tasca, D. S., Boyd, R. W. & Padgett, M. J. EPR-based ghost imaging using a single-photon-sensitive camera. *New J. Phys.* **15**, 073032 (2013).
10. Wiseman, H. M. & Mølmer, K. Induced coherence with and without induced emission. *Phys. Lett. A* **270**, 245–248 (2000).
11. Elitzur, A. C. & Vaidman, L. Quantum mechanical interaction-free measurements. *Found. Phys.* **23**, 987–997 (1993).
12. Kwiat, P., Weinfurter, H., Herzog, T., Zeilinger, A. & Kasevich, M. A. Interaction-free measurement. *Phys. Rev. Lett.* **74**, 4763–4766 (1995).
13. Horne, M. in *Experimental Metaphysics* Vol. 1. (eds Cohen, R. S., Horne, M. & Stachel, J.) 109–119 (Kluwer Academic, 1997).
14. Howell, J. C., Bennink, R. S., Bentley, S. J. & Boyd, R. W. Realization of the Einstein-Podolsky-Rosen paradox using momentum- and position-entangled photons from spontaneous parametric down conversion. *Phys. Rev. Lett.* **92**, 210403 (2004).
15. Walborn, S. P., Monken, C. H., Pádua, S. & Souto Ribeiro, P. H. Spatial correlations in parametric down-conversion. *Phys. Rep.* **495**, 87–139 (2010).
16. Tasca, D. S., Walborn, S. P., Souto Ribeiro, P. H., Toscano, F. & Pellat-Finiet, P. Propagation of transverse intensity correlations of a two-photon state. *Phys. Rev. A* **79**, 033801 (2009).
17. Horne, M. A., Shimony, A. & Zeilinger, A. Two particle interferometry. *Phys. Rev. Lett.* **62**, 2209–2212 (1989); Two particle interferometry. *Nature* **347**, 429–430 (1990).
18. Ribeiro, P. H. S., Pádua, S., Machado da Silva, J. C. & Barbosa, G. A. Controlling the degree of visibility of Young's fringes with photon coincidence measurements. *Phys. Rev. A* **49**, 4176–4179 (1994).
19. Abouraddy, A. F., Stone, P. R., Sergienko, A. V., Saleh, B. E. A. & Teich, M. C. Entangled-photon imaging of a pure phase object. *Phys. Rev. Lett.* **93**, 213903 (2004).
20. Zou, X. Y., Grayson, T. P. & Mandel, L. Observation of quantum interference effects in the frequency domain. *Phys. Rev. Lett.* **69**, 3041–3044 (1992).

21. Grayson, T. P. & Barbosa, G. A. Spatial properties of spontaneous parametric down-conversion and their effect on induced coherence without induced emission. *Phys. Rev. A* **49**, 2948–2961 (1994).

Acknowledgements We thank M. Horne for reading the manuscript, clarifying suggestions and many discussions, P. Enigl for designing the figures for the objects, D. Greenberger and S. von Egan-Krieger for discussions, and C. Schaeff for equipment loans. Microfabrication was carried out at the Center for Micro- and Nanostructures (ZMNS) of the Vienna University of Technology. We acknowledge D. Ristanic for assistance with cryogenic Si etching and M. Schinnerl for contact mask production. G.B.L. was funded by the Austrian Academy of Sciences (ÖAW) through a fellowship from the Vienna Center for Science and Technology (VCQ). S.R. is funded by an EU Marie Curie Fellowship (PIOF-GA-2012-329851). This project was supported by ÖAW, the European Research Council (ERC Advanced grant no. 227844 ‘QIT4QAD’), and SIQS

grant no. 600645 EU-FP7-ICT), and the Austrian Science Fund (FWF) with SFB F40 (FOQUS) and W1210-2 (CoQus).

Author Contributions A.Z. initiated this research. G.B.L., V.B., R.L., S.R. and A.Z. designed the experiment. G.B.L., V.B. and R.L. carried out the experiment. G.D.C. fabricated the silicon and silica phase masks. All authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.B.L. (gabriela.barreto.lemos@univie.ac.at) and A.Z. (anton.zeilinger@univie.ac.at).

Carbon–carbon bond cleavage and rearrangement of benzene by a trinuclear titanium hydride

Shaowei Hu¹, Takanori Shima^{1,2} & Zhaomin Hou^{1,2}

The cleavage of carbon–carbon (C–C) bonds by transition metals is of great interest, especially as this transformation can be used to produce fuels and other industrially important chemicals from natural resources such as petroleum and biomass. Carbon–carbon bonds are quite stable and are consequently unreactive under many reaction conditions. In the industrial naphtha hydrocracking process, the aromatic carbon skeleton of benzene can be transformed to methylcyclopentane and acyclic saturated hydrocarbons through C–C bond cleavage and rearrangement on the surfaces of solid catalysts^{1–6}. However, these chemical transformations usually require high temperatures and are fairly non-selective. Microorganisms can degrade aromatic compounds under ambient conditions, but the mechanistic details are not known and are difficult to mimic⁷. Several transition metal complexes have been reported to cleave C–C bonds in a selective fashion in special circumstances, such as relief of ring strain, formation of an aromatic system, chelation-assisted cyclometallation and β -carbon elimination^{8–15}. However, the cleavage of benzene by a transition metal complex has not been reported^{16–19}. Here we report the C–C bond cleavage and rearrangement of benzene by a trinuclear titanium polyhydride complex. The benzene ring is transformed sequentially to a methylcyclopentenyl and a 2-methylpentenyl species through the cleavage of the aromatic carbon skeleton at the multi-titanium sites. Our results suggest that multinuclear titanium hydrides could serve as a unique platform for the activation of aromatic molecules, and may facilitate the design of new catalysts for the transformation of inactive aromatics.

The investigation of the reactivity of multimetallic complexes with benzene is of great interest and importance, in view of the fact that the activation of aromatic compounds on industrial solid catalysts might rely on the cooperation of multiple metal sites^{1–6,20,21}. In our recent studies on multinuclear rare-earth and transition metal hydride complexes, we found that the multiple metal–hydride sites could have unique synergistic effects on the activation of small molecules such as CO, CO₂, H₂ and N₂ (refs 22–26). In particular, the unusually high reactivity of the trinuclear titanium heptahydride complex [(C₅Me₄SiMe₃)Ti]₃(μ -H)₆ (**1**)²⁶ motivated us to examine carefully whether the hydride cluster could activate other chemically inactive molecules such as benzene.

When a benzene solution of **1** was left to stand at room temperature (22 °C) under an argon atmosphere for a few days, a methylcyclopentenyl complex [(C₅Me₄SiMe₃)Ti]₃(C₅H₄Me)(μ -H)₄ (**2**) was formed as dark green crystals. Complex **2** could be obtained almost quantitatively, when a benzene solution of **1** was kept at 40 °C for 36 h (Fig. 1a). An X-ray diffraction study revealed that **2** contains one methylcyclopentenyl (MeC₅H₄) moiety, three (C₅Me₄SiMe₃)Ti units and four hydride ligands (Fig. 1a). The methylcyclopentenyl moiety formally bears three negative charges, one being distributed among C39, C40 and C41 to form an allylic moiety and two being placed on C37 and C38. Four (C37, C38, C39, C41) of the five ring carbon atoms in the [MeC₅H₄]^{3–} unit are μ - η^4 -bonded to the Ti₃ triangle.

Remarkably, here a benzene molecule was partly hydrogenated and ring-contracted to a five-membered ring species, [MeC₅H₄]. This reaction

required breaking a robust aromatic C–C bond of benzene and making a new C–C bond. In this transformation, three of the seven hydride ligands in **1** were consumed, one being transferred to benzene and two released as H₂ by donating two electrons, thus affording the trianionic [MeC₅H₄]^{3–} species. The formal oxidation state of the Ti atoms in **1** and **2** remained unchanged; both complexes formally contain one Ti(IV) and two Ti(III) atoms²⁶.

When ¹³C-enriched benzene ¹³C₆H₆ was used to react with **1** under similar conditions, the corresponding ¹³C-enriched methylcyclopentenyl complex, [(C₅Me₄SiMe₃)Ti]₃(¹³C₅H₄¹³CH₃)(μ -H)₄ (**2**-¹³C), was obtained (Supplementary Information), confirming that the methylcyclopentenyl unit in **2** was indeed derived from benzene. The reaction of **1** with benzene-*d*₆ (C₆D₆) afforded [(C₅Me₄SiMe₃)Ti]₃[C₅Z₄CZ₃](μ -Z)₄ (Z = H_{0.7–0.5}D_{0.3–0.5}) (**2**-HD), in which H and D atoms are randomly distributed between the methylcyclopentenyl unit and the titanium framework as shown by ¹H and ²H NMR spectroscopy.

To isolate a possible reaction intermediate, we kept a benzene solution of **1** at a lower temperature (10 °C) for about two days. Evaporation of the solvent under vacuum yielded dark-green solids, which appeared to be a mixture of **2** (25%) and a new compound, [(C₅Me₄SiMe₃)Ti]₃(μ - η^6 -C₆H₆)(μ -H)₃ (**4**) (75%), as shown by ¹H NMR spectroscopy. Although complete separation of **2** and **4** was difficult, a single crystal of **4** suitable for X-ray diffraction studies was obtained by recrystallization in tetrahydrofuran (THF). We found that **4** is formally composed of a tetra-anionic [C₆H₆]^{4–} moiety and a tetra-cationic titanium trihydride unit, [Cp'₃Ti₃H₃]⁴⁺ (Fig. 1a). The [C₆H₆] unit is μ - η^6 -bonded to the Ti₃ framework. The ¹H NMR spectrum of **4** exhibited a singlet at a chemical shift of δ = 4.88 p.p.m. (tetramethylsilane reference) for the six protons of the [C₆H₆] unit in a temperature range of 22 °C to –80 °C, showing that the structure is highly fluxional.

Complex **4** decomposed slowly to unidentified products at room temperature. The direct transformation of **4** to **2** was not observed. When **4** was exposed to an atmosphere of H₂ in THF-*d*₈, the immediate formation of another new compound assignable to [(C₅Me₄SiMe₃)Ti]₃(C₆H₇)(μ -H)₄ (**3**) was observed (Supplementary Information; see also **6** in Fig. 2, and below). In this reaction, the [C₆H₆]^{4–} unit in **4** was formally hydrogenated (or protonated) to [C₆H₇]^{3–}.

When kept at 40 °C for about 3 h, **3** was transformed quantitatively to the methylcyclopentenyl [MeC₅H₄]^{3–} complex **2** through ring contraction of the [C₆H₇]^{3–} unit (Fig. 1a). This transformation occurred even at room temperature, albeit slower. Kinetic studies on the transformation of **3** to **2** revealed the activation parameters of ΔH^\ddagger = 21.7(6) kcal mol^{–1} and ΔS^\ddagger = –4.7(19) entropy units (Supplementary Information). The entropy of activation is in agreement with a unimolecular process. In the absence of H₂, **3** rapidly changed to **4** by release of H₂, demonstrating that **3** and **4** are easily interconvertible through hydrogenation/dehydrogenation.

To gain more information on the reaction process, we monitored the reaction of **1** with benzene-*d*₆ (C₆D₆) at 22 °C by ¹H NMR spectroscopy. With the decrease of **1**, gradual formation of **2**-HD was observed (Fig. 1b). In this process, simultaneous formation of two species assignable

¹Advanced Catalysis Research Group, RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. ²Organometallic Chemistry Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

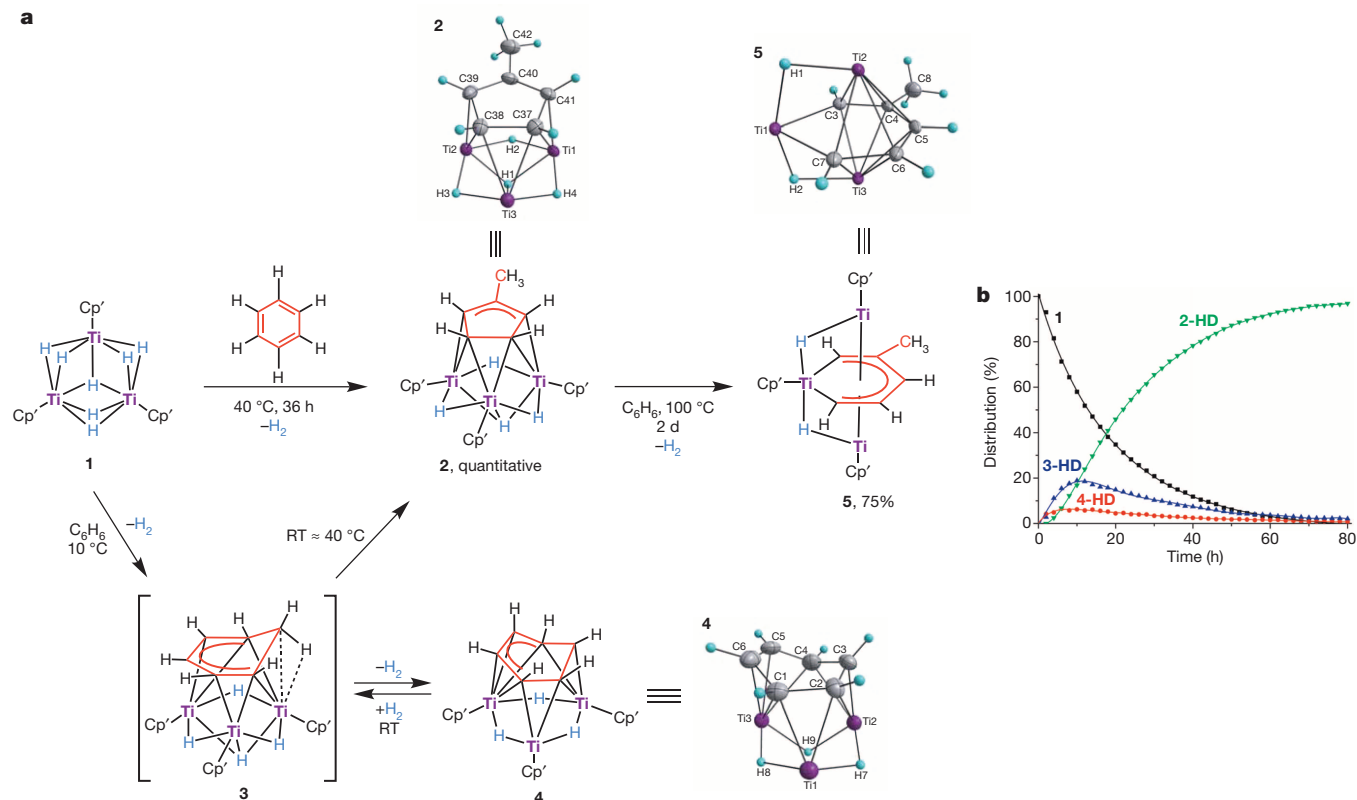


Figure 1 | Reactions of a trinuclear titanium heptahydride complex (1) with benzene and benzene- d_6 . **a**, Reaction of 1 with benzene ($\text{Cp}' = \text{C}_5\text{Me}_4\text{SiMe}_3$). The X-ray core structures of 2, 4 and 5 are shown as ball and stick diagrams

to 3-HD and 4-HD was also observed. After 78 h, 3-HD, 4-HD and 1 all disappeared to yield 2-HD almost quantitatively.

The above experimental results suggest that 3 may be first formed by the reductive hydrogenation of benzene by the heptahydride complex 1 through hydride transfer and H_2 elimination (Fig. 1a). There is an equilibrium between 3 and 4 through dehydrogenation and rehydrogenation. The isomerization (ring contraction) of the $[\text{C}_6\text{H}_7]$ unit in 3 would finally yield the methylcyclopentenyl complex 2.

When 2 was heated at 100°C for two days in a sealed Schlenk tube, insertion of a Ti atom into a C–C bond of the methylcyclopentenyl ring took place, yielding a titanacycle product, $[(\text{C}_5\text{Me}_4\text{SiMe}_3)_3\text{Ti}_3][\mu-\eta^2, \eta^5, \eta^5\text{-CHC}(\text{Me})(\text{CH})_3](\mu\text{-H})_2$ (5), as a major product (about 75%) (Fig. 1a). This transformation was accompanied by liberation of H_2 . Heating 2- ^{13}C under the same conditions (in benzene) afforded the corresponding

(Cp' ligands omitted for clarity). Me, methyl. **b**, Conversion versus time for the reaction of 1 with benzene- d_6 at 22°C . The solid lines are interpolations.

^{13}C -enriched complex, 5- ^{13}C , suggesting that this rearrangement should be a unimolecular process. Complex 5 could formally be viewed as a combination of a penta-anionic 2-methylpentenyl group, $[\text{CHC}(\text{Me})(\text{CH})_3]^{5-}$, and a penta-cationic titanium dihydride unit, $[(\text{C}_5\text{Me}_4\text{SiMe}_3)_3\text{Ti}_3\text{H}_2]^{5+}$. An X-ray diffraction study established that one $(\text{C}_5\text{Me}_4\text{SiMe}_3)_3\text{Ti}$ unit is incorporated into the 2-methylpentenyl framework to form a planar, six-membered titanacycle. Each side of the metallacycle plane is η^5 -bonded to a $(\text{C}_5\text{Me}_4\text{SiMe}_3)_3\text{Ti}$ unit, leading to formation of a double stacker structure. There are two hydride ligands in 5, each bridging the Ti atom in the metallacycle and one Ti atom outside of the metallacycle.

For comparison, we also examined the reaction of 1 with toluene. When a toluene solution of 1 was kept at 20°C for 42 h, the partial hydrogenation of toluene took place to give $[(\text{C}_5\text{Me}_4\text{SiMe}_3)_3\text{Ti}_3](\text{C}_6\text{H}_6\text{Me})(\mu\text{-H})_4$ (6), which contains a $[\text{C}_6\text{H}_6\text{Me}]^{3-}$ unit and could be viewed as an analogue

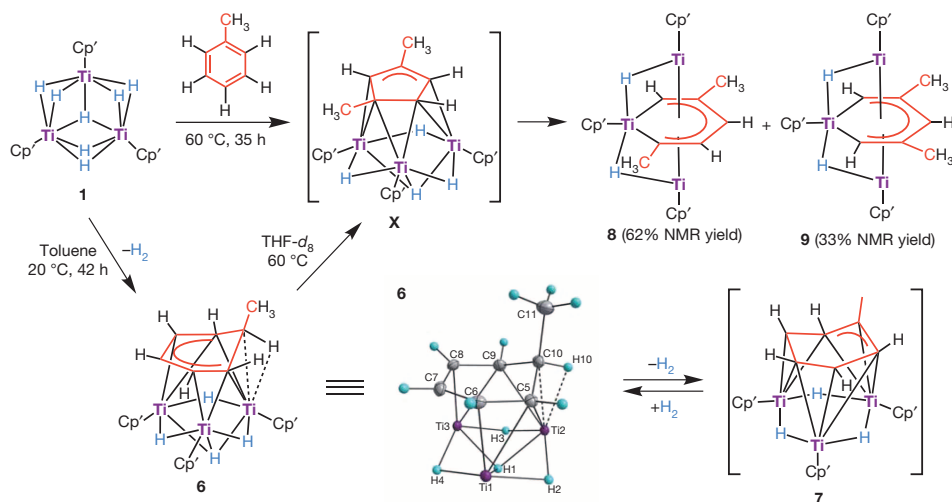


Figure 2 | Reaction of complex 1 with toluene. The X-ray core structure of 6 is shown as a ball and stick diagram (Cp' ligands omitted for clarity).

of **3** formed in the reaction of **1** with benzene (Fig. 2; see also Fig. 1a). A single crystal of **6** suitable for X-ray diffraction studies was obtained by recrystallization under an atmosphere of H₂ in hexane. It was revealed that there is an agostic interaction between a C–H bond in the C₆H₆Me unit, which is formed by hydrogenation of toluene at the *ipso*-position, and a Ti atom (Ti2) (Ti2–C10: 2.326(3) Å; Ti2–H10: 2.11(3) Å). There are four hydride ligands in **6**, one capping the Ti₃ triangle in a μ_3 -fashion and three bridging the Ti···Ti sides in a μ_2 -mode, analogously to complex **2** (Figs 2 and 1a). In agreement with the solid structure, the *ipso*-proton in the [C₆H₆Me] unit in **6** exhibited a singlet at a high field $\delta = -2.23$ p.p.m. in the ¹H NMR spectrum, showing that an agostic Ti···H interaction remained in solution. The ¹³C-enriched analogue **6**-¹³C, prepared by the reaction of **1** with ¹³C-enriched toluene (¹³C₆H₅¹³CH₃), showed a doublet at $\delta = -2.23$ p.p.m. with a relatively small ¹³C–¹H coupling constant ($J_{C-H} = 108$ Hz) in the ¹H NMR spectrum, thus providing further proof for the presence of an agostic Ti···H interaction (for comparison, normally $J_{C-H} = 120$ –130 Hz; ref. 27).

In the absence of H₂, the dehydrogenation of **6** took place rapidly to give [(C₅Me₄SiMe₃)Ti]₃(C₆H₅Me)(μ -H)₃ (**7**) (Fig. 2), an analogue of **4** formed in the case of benzene. When **6** (or **1** plus toluene) was heated in a THF-*d*₈ solution at 60 °C for 36 h, the six-membered ring metallocycle products [(C₅Me₄SiMe₃)Ti]₃[μ - η^2 , η^5 , η^5 -CHC(Me)(CH)₂C(Me)](μ -H)₂ (**8**) (62% NMR yield) and [(C₅Me₄SiMe₃)Ti]₃[μ - η^2 , η^5 , η^5 -CHC(Me)(CH)C(Me)CH](μ -H)₂ (**9**) (33% NMR yield), which are analogues of **5** formed in the case of benzene, were obtained. The formation of the two regioisomers **8** and **9** suggests that a dimethylcyclopentenyl species such as **X** (an analogue of **2**; Fig. 2) should be an intermediate in this reaction. But in contrast to the reaction of benzene, such a five-membered ring intermediate was not observed in the case of toluene (at either 60 °C or lower temperatures), possibly because of its instability caused by the steric hindrance of the methyl groups on the cyclopentenyl ring.

Obviously, the reactions of the hydride cluster **1** with benzene and toluene involved various events, including C = C double-bond hydrometallation, H₂ release, and C–H and C–C bond cleavage and formation. It is certainly due to the cooperation of the multiple Ti–H sites that all of these different events could be accomplished at once, although the mechanistic details are subject to further studies. It is also worth noting that the C–C bond cleavage of cyclic alkenes such as cyclopentadiene and cyclohexene has been successfully achieved by using multimetallic ruthenium hydride and carbonyl complexes^{28,29}, but the breaking of a benzene ring did not take place under similar conditions³⁰. The present sequential transformation of benzene to **2** and **5** by the hydride cluster **1** also stands in contrast with the hydrocracking reaction of a benzene ring on solid catalysts, which gave a mixture of methylcyclopentane and acyclic saturated hydrocarbons^{1–6}. This work thus demonstrates that multimetallic titanium hydrides such as **1** can serve as a unique platform for the activation of aromatic molecules, offering new opportunities for the transformation and functionalization of inactive aromatics.

METHODS SUMMARY

All reactions were performed under an argon atmosphere. Complex **1** was prepared according to literature procedures²⁶ and stored in a glovebox.

Reaction of **1 with benzene.** A benzene solution (5.0 ml) of **1** (151 mg, 0.21 mmol) in a 30 ml Schlenk tube was stirred at 40 °C for 36 h. Removal of the solvent under vacuum gave [(C₅Me₄SiMe₃)Ti]₃(C₆H₄Me)(μ -H)₄ (**2**) (163 mg, 98%) as a dark-green solid. Recrystallization from hexane at –33 °C afforded dark-green crystals (121 mg, 0.15 mmol, 71%) suitable for X-ray diffraction.

Received 1 February; accepted 27 June 2014.

- Jones, D. S. J. & Pujadó, P. R. *Handbook of Petroleum Processing* (Springer, 2006).
- Weitkamp, J. *Handbook of Heterogeneous Catalysis* Vol. 1, 2nd edn (eds Ertl, G., Knözinger, H., Schüth, F. & Weitkamp, J.) Ch. 14.2 (Wiley-VCH, 2008).
- Watanabe, R., Suzuki, T. & Okuhara, T. Skeletal isomerization of alkanes and hydroisomerization of benzene over solid strong acids and their bifunctional catalysts. *Catal. Today* **66**, 123–130 (2001).
- McVicker, G. B. et al. Selective ring opening of naphthenic molecules. *J. Catal.* **210**, 137–148 (2002).

- Benitez, V. M., Grau, J. M., Yori, J. C., Pieck, C. L. & Vera, C. R. Hydroisomerization of benzene-containing paraffinic feedstocks over Pt/WO₃–ZrO₂ catalysts. *Energy Fuels* **20**, 1791–1798 (2006).
- Kazakov, M. O. et al. Hydroisomerization of benzene-containing gasoline fractions on a Pt/SO₄^{2–}–ZrO₂–Al₂O₃ catalyst: I. Effect of chemical composition on the phase state and texture characteristics of SO₄^{2–}–ZrO₂–Al₂O₃ supports. *Kinet. Catal.* **51**, 438–443 (2010).
- Bugg, T. D. H. & Winfield, C. J. Enzymatic cleavage of aromatic rings: mechanistic aspects of the catechol dioxygenases and later enzymes of bacterial oxidative cleavage pathways. *Nat. Prod. Rep.* **15**, 513–530 (1998).
- Bishop, K. C. Transition metal catalysed rearrangements of small ring organic molecules. *Chem. Rev.* **76**, 461–486 (1976).
- Crabtree, R. H. The organometallic chemistry of alkanes. *Chem. Rev.* **85**, 245–269 (1985).
- Jones, W. D. The fall of the C–C bond. *Nature* **364**, 676–677 (1993).
- Rybtchinski, B. & Milstein, D. Metal insertion into C–C bonds in solution. *Angew. Chem. Int. Ed.* **38**, 870–883 (1999).
- Jun, C. Transition metal-catalysed carbon–carbon bond activation. *Chem. Soc. Rev.* **33**, 610–618 (2004).
- Murakami, M. & Matsuda, T. Metal-catalysed cleavage of carbon–carbon bonds. *Chem. Commun.* **47**, 1100–1105 (2011).
- Ruhland, K. Transition-metal-mediated cleavage and activation of C–C single bonds. *Eur. J. Org. Chem.* **2012**, 2683–2706 (2012).
- Takao, T. & Suzuki, H. Skeletal rearrangement of hydrocarbyl ligands on a triruthenium core induced by chemical oxidation. *Coord. Chem. Rev.* **256**, 695–708 (2012).
- Sattler, A. & Parkin, G. Cleaving carbon–carbon bonds by inserting tungsten into unstrained aromatic rings. *Nature* **463**, 523–526 (2010).
- Kira, M., Ishida, S., Iwamoto, T. & Kabuto, C. Excited-state reactions of an isolable silylene with aromatic compounds. *J. Am. Chem. Soc.* **124**, 3830–3831 (2002).
- Ellis, D., McKay, D., Macgregor, S. A., Rosair, G. M. & Welch, A. J. Room-temperature C–C bond cleavage of an arene by a metallocarborane. *Angew. Chem. Int. Ed.* **49**, 4943–4945 (2010).
- Szyszkowski, B., Latos-Grażyński, L. & Sztörenberg, L. A facile palladium-mediated contraction of benzene to cyclopentadiene: transformations of palladium(II) *p*-benzoporphyryr. *Angew. Chem. Int. Ed.* **50**, 6587–6591 (2011).
- Van Hove, M. A., Lin, R. F. & Somorjai, G. A. Surface structure of coadsorbed benzene and carbon monoxide on the rhodium(III) single crystal analysed with low energy electron diffraction intensities. *J. Am. Chem. Soc.* **108**, 2532–2537 (1986).
- Dyson, P. J. Arene hydrogenation by homogeneous catalysts: fact or fiction? *Dalton Trans.* 2964–2974 (2003).
- Tardif, O., Hashizume, D. & Hou, Z. Hydrogenation of carbon dioxide and aryl isocyanates by a tetranuclear tetrahydrido yttrium complex. isolation, structures, and CO₂ insertion reactions of methylene diolate and μ_3 -oxo yttrium complexes. *J. Am. Chem. Soc.* **126**, 8080–8081 (2004).
- Shima, T. & Hou, Z. Hydrogenation of carbon monoxide by tetranuclear rare earth metal polyhydrido complexes. selective formation of ethylene and isolation of well-defined polyoxo rare earth metal clusters. *J. Am. Chem. Soc.* **128**, 8124–8125 (2006).
- Nishiura, M. & Hou, Z. Novel polymerization catalysts and hydride clusters from rare-earth metal dialkyls. *Nature Chem.* **2**, 257–268 (2010).
- Shima, T. et al. Molecular heterometallic hydride clusters composed of rare-earth and d-transition metals. *Nature Chem.* **3**, 814–820 (2011).
- Shima, T. et al. Dinitrogen cleavage and hydrogenation by a trinuclear titanium polyhydride complex. *Science* **340**, 1549–1552 (2013).
- Brookhart, M. & Green, M. L. H. Carbon–hydrogen–transition metal bonds. *J. Organomet. Chem.* **250**, 395–408 (1983).
- Suzuki, H., Takaya, Y., Takemori, T. & Tanaka, M. Selective carbon–carbon bond cleavage of cyclopentadiene on a trinuclear ruthenium pentahydride complex. *J. Am. Chem. Soc.* **116**, 10779–10780 (1994).
- Brown, D. B. et al. Cluster-mediated ring contraction: synthesis and characterisation of [Ru₆(μ_3 -H)(μ_4 - η^2 -CO)₂(CO)₁₃(η^5 -C₅H₄Me)] and [Ru₆(μ_3 -H)(μ_4 - η^2 -CO)₂(CO)₁₃(η^5 -C₅H₃C₃H₆)]. *J. Chem. Soc. Dalton Trans.* 1909–1914 (1997).
- Takao, T. et al. Synthesis and property of diruthenium complexes containing bridging cyclic diene ligands and the reaction of diruthenium tetrahydrido complex with benzene forming a μ - η^2 : η^2 -cyclohexadiene complex via partial hydrogenation on a Ru₂ centre. *Organometallics* **30**, 5057–5067 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a Grant-in-Aid for Young Scientists (B) (no. 26810041), a Grant-in-Aid for Scientific Research (C) (no. 26410082) and a Grant-in-Aid for Scientific Research (S) (no. 26220802) from JSPS, and an Incentive Research Grant from RIKEN. We thank J. Cheng for help with X-ray structure analyses, and A. Karube for conducting elemental analyses.

Author Contributions Z.H., S.H. and T.S. had the idea for and designed the experiments. S.H. and T.S. conducted the experiments. Z.H. and S.H. wrote the manuscript. All authors participated in data analyses and discussions. Z.H. directed the project.

Author Information X-ray crystallographic coordinates of **2**, **4**, **5** and **6** have been deposited at the Cambridge Crystallographic Database under accession numbers 981670–981673. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.H. (houz@riken.jp).

Contrasting responses of mean and extreme snowfall to climate change

Paul A. O’Gorman¹

Snowfall is an important element of the climate system, and one that is expected to change in a warming climate^{1–4}. Both mean snowfall and the intensity distribution of snowfall are important, with heavy snowfall events having particularly large economic and human impacts^{5–7}. Simulations with climate models indicate that annual mean snowfall declines with warming in most regions but increases in regions with very low surface temperatures^{3,4}. The response of heavy snowfall events to a changing climate, however, is unclear. Here I show that in simulations with climate models under a scenario of high emissions of greenhouse gases, by the late twenty-first century there are smaller fractional changes in the intensities of daily snowfall extremes than in mean snowfall over many Northern Hemisphere land regions. For example, for monthly climatological temperatures just below freezing and surface elevations below 1,000 metres, the 99.99th percentile of daily snowfall decreases by 8% in the multimodel median, compared to a 65% reduction in mean snowfall. Both mean and extreme snowfall must decrease for a sufficiently large warming, but the climatological temperature above which snowfall extremes decrease with warming in the simulations is as high as -9°C , compared to -14°C for mean snowfall. These results are supported by a physically based theory that is consistent with the observed rain–snow transition. According to the theory, snowfall extremes occur near an optimal temperature that is insensitive to climate warming, and this results in smaller fractional changes for higher percentiles of daily snowfall. The simulated changes in snowfall that I find would influence surface snow and its hazards; these changes also suggest that it may be difficult to detect a regional climate-change signal in snowfall extremes.

Extremes of daily precipitation (including liquid and solid precipitation) are found to increase in intensity with climate warming in observations and simulations^{8–10}, and this is physically consistent with greater saturation-specific humidities in a warmer atmosphere^{11–13}. However, little is known about the physical basis for changes in daily snowfall extremes, their past changes on a global or hemispheric scale, or how they change in global-climate-model simulations. Regional observational studies show large interdecadal variations in measures of snowfall extremes^{14,15}, but long-term trends remain unclear. Extremes of seasonal mean snowfall have been studied previously^{16,17}, but daily snowfall extremes may respond differently¹⁴. Physically, we would expect heavy snowfall events to occur in a relatively narrow range of temperatures below the rain–snow transition; at much lower temperatures it is not ‘too cold to snow’ but low saturation-specific humidities make heavy snowfall unlikely. However, it is not clear what this means for the response to climate change, and previous studies have differed in their findings as to whether heavy snowfall events are predominantly associated with anomalously cold or warm years (or seasons) in the present climate^{14,18}. Variability of daily temperatures is another factor that must be taken into account, and cold extremes are expected to persist to some extent in a warming climate¹⁹.

I show here, using simulations and a physically based theory, that snowfall extremes respond more weakly to climate warming than does mean snowfall in many regions. The simulations were performed with 20 climate models and were taken from the World Climate Research

Programme’s Coupled Model Intercomparison Project phase 5 (CMIP5), which is the first phase of the project to archive daily snowfall (Methods). Climate change was calculated as the difference between the historical simulations (1981–2000; the control climate) and the representative concentration pathway (RCP) 8.5 simulations (2081–2100; the warm climate). The snowfall variable is accumulated daily, includes all solid precipitation at the surface and is expressed in liquid-water equivalent per day (extremes of snowfall depth are discussed in the Methods). For simplicity, only Northern Hemisphere land was considered, and results are presented as the multimodel median of the ratio of snowfall rates in the warm versus the control climate.

Daily snowfall extremes were first measured by their 20-year return values, calculated by fitting the generalized extreme value distribution to the time series of annual maximum daily snowfall in each grid box (Methods). Compared to observational estimates of snowfall, the control simulations capture the magnitudes and many of the features of mean and extreme snowfall, with some regional biases (Extended Data Figs 1 and 2). Climate warming in the simulations causes widespread decreases in mean snowfall at middle latitudes (Fig. 1a), consistent with previous studies^{3,4}. In contrast, the snowfall extremes have a relatively muted

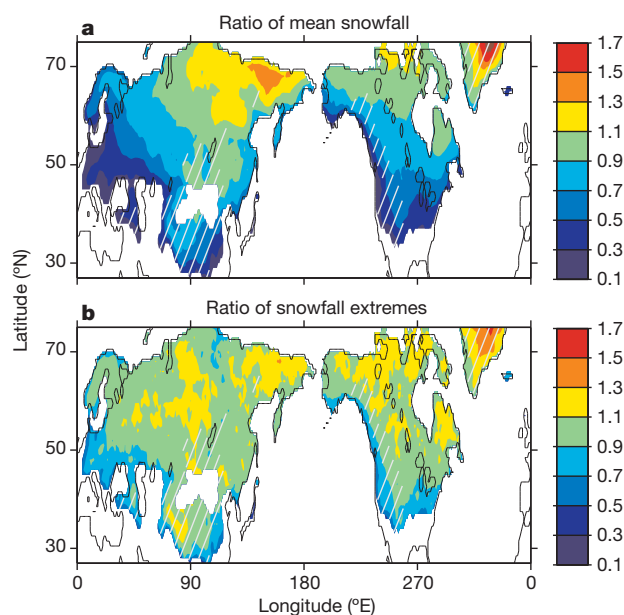


Figure 1 | Ratios of snowfall for the warm climate compared with the control climate. **a, b,** Multimodel-median ratios (colour scale) of mean snowfall (**a**) and daily snowfall extremes as measured by their 20-year return values (**b**). The 20-year return values were estimated using a fit of the generalized extreme value distribution to the annual-maximum time series. Ratios are only shown for land grid boxes where the multimodel-median of mean snowfall is greater than 5 cm per year in the control climate. White hatching denotes regions with surface elevations above 1,000 m that are not included in Figs 2–4.

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

response, with substantially smaller fractional changes than for mean snowfall in many regions (Fig. 1b).

Snowfall statistics and their changes are expected to be strongly dependent on the climatological temperature, which varies by month and region. To quantify this dependence, I next analysed the changes in snowfall as a function of the climatological monthly surface air temperature in the control climate. Daily snowfall rates were aggregated in 5 °C bins with centres from −22.5 °C to 12.5 °C according to the climatological monthly surface air temperature in the control climate for each grid box and day. Snowfall extremes were calculated as high percentiles of the daily snowfall rates in each temperature bin, including days with no snowfall. Both mean snowfall and snowfall extremes in the different temperature bins are in good agreement with observational estimates (Extended Data Fig. 3). The response to climate change is first presented for surface elevations below 1,000 m (Extended Data Fig. 4). Fractional decreases are greater for mean snowfall than for snowfall extremes for much of the temperature range considered here (Fig. 2a), which demonstrates the contrasting responses of mean and extreme snowfall even when monthly variations in climatological temperature are controlled for. For the temperature bin centred around −2.5 °C, mean snowfall decreases by 65% in the multimodel median, whereas the 99.99th percentile of snowfall decreases by only 8%. Changes in snowfall extremes transition from positive to negative at control-climate temperatures as high as −9 °C, whereas the corresponding temperature for mean snowfall is −14 °C. Furthermore, the difference in behaviour between mean and extremes is greater the higher the percentile of snowfall considered (Fig. 2a), and this difference is robust across different climate models (Extended Data Fig. 5).

I next present a simple theory that accounts for the main features of the response of snowfall extremes to climate change. The theory does not include the response of mean snowfall, but this has been explained previously in terms of changes in mean precipitation and temperature^{3,4}. Surface precipitation type depends on the vertical temperature profile

of the lower troposphere²⁰, but to first order it may be related to surface air temperature^{21,22}. The daily snowfall rate s in the theory is related to the daily precipitation rate p by $s = f(T)p$, where T is the daily surface air temperature, and $f(T)$ is the snowfall fraction (the fraction of precipitation that falls as snow at a given temperature T). The $f(T)$ diagnosed from the simulations shows a sharp decline near freezing (Fig. 3), and this is comparable to what is found in observations (Extended Data Fig. 6). As expected given modest changes in lapse rates (the rates of decrease of temperature with height), $f(T)$ is almost exactly the same in the control and warm climates (Fig. 3).

The daily precipitation rate in the theory is assumed to have a simple dependence on surface air temperature according to $p = e^{\beta T} \hat{p}$, where $\beta = 0.06 \text{ °C}^{-1}$ is a representative thermodynamic rate of increase of extratropical precipitation extremes with respect to surface temperature related to changes in saturation-specific humidity¹². The normalized precipitation variable \hat{p} may be thought of as a dynamic variable closely related to upward motion in the atmosphere; it is assumed to follow a gamma distribution on wet days with scale parameter γ^{-1} and shape parameter k . The fraction of wet days is denoted w . The temperature T is assumed to be normally distributed with mean \bar{T} and standard deviation σ , and \hat{p} and T are taken to be independent.

With these assumptions, asymptotic methods were used to evaluate the integrals over temperature and \hat{p} involved in the calculation of high percentiles of snowfall (Methods). The reciprocal of the temperature dependence of the snowfall rate is denoted $h(T) = e^{-\beta T} f(T)^{-1}$, and the asymptotics show that the behaviour of snowfall extremes is dominated by the temperature T_m at which $h(T)$ reaches a minimum (roughly −2 °C in the simulations and observations). T_m is the optimal temperature for snowfall extremes in the theory, and it arises because of the competition between increasing saturation-specific humidity and decreasing snowfall fraction with increasing temperature. The result is that the q th percentile of snowfall s_q is given by:

$$(\gamma s_q h_m)^{\frac{3}{2}-k} e^{\gamma s_q h_m} = \frac{w}{\sigma \left(1 - \frac{q}{100}\right) \Gamma(k)} \sqrt{\frac{h_m}{h_m''}} e^{-\frac{(\bar{T} - T_m)^2}{2\sigma^2}} \quad (1)$$

which is valid asymptotically for large s_q values, where Γ is the gamma function, h_m is h evaluated at T_m , and h_m'' is the second derivative of h at T_m . For a change in mean temperature of $\delta \bar{T}$ and assuming negligible changes in all other parameters, the change in snowfall extremes, δs_q , is given by:

$$\delta s_q = -\frac{\delta \bar{T}}{\sigma^2 \gamma h_m} \left(\bar{T} + \frac{\delta \bar{T}}{2} - T_m \right) \quad (2)$$

as shown in the Methods.

According to equation (2), δs_q transitions from positive to negative at a mean temperature in the control climate of $T_m - \delta \bar{T}/2$ (roughly equal to −6 °C in the simulations), and it is proportional to $1/(\gamma h_m)$,

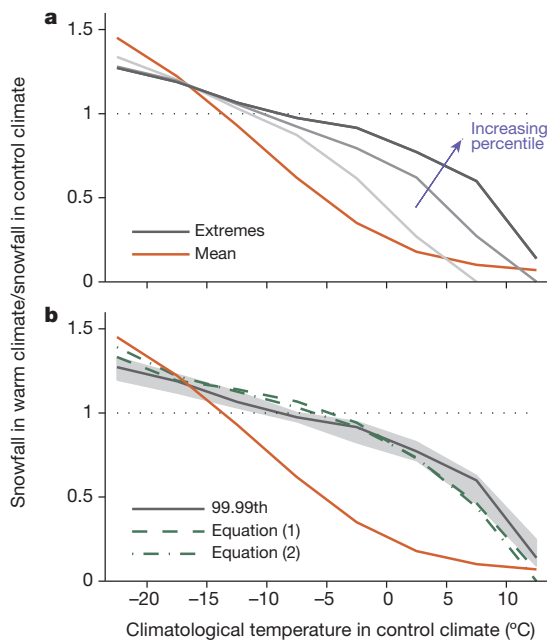


Figure 2 | Ratios of snowfall for the warm climate compared with the control climate as a function of climatological monthly surface air temperature in the control climate. Multimodel-median ratios of mean snowfall (red) in both panels. **a**, Multimodel-median ratios of the 99th, 99.9th and 99.99th percentiles of daily snowfall in increasing order from light to dark grey. **b**, Multimodel-median ratio of the 99.99th percentile of daily snowfall (grey line; shading shows the interquartile range), and the same ratio calculated from theory according to equation (1) (green dashed) and equation (2) (green dashed-dotted). Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.

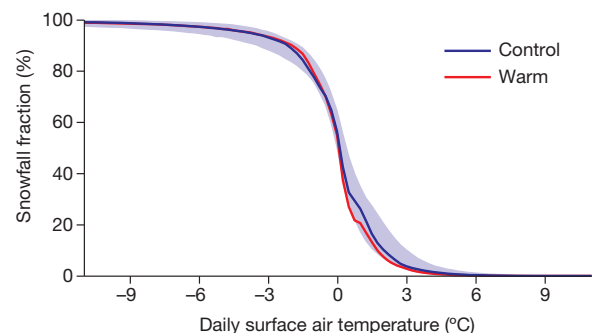


Figure 3 | Daily snowfall fraction as a function of daily surface air temperature. The multimodel-median snowfall fraction is shown for the control climate (blue line; shading shows the interquartile range) and the warm climate (red line). It is calculated in each model and for each climate as the ratio of mean snowfall to mean precipitation in daily temperature bins of width 0.25 °C. Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.

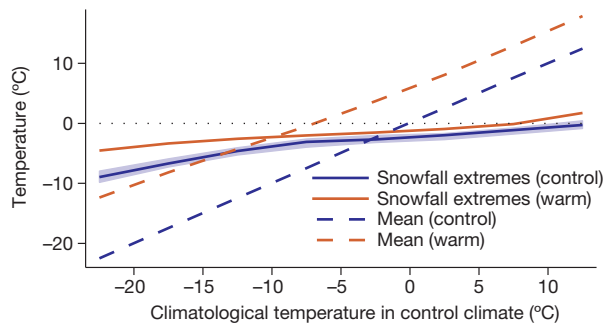


Figure 4 | Multimodel-median surface air temperatures at which snowfall extremes occur as a function of climatological monthly surface air temperature in the control climate. For each control-climate temperature bin, surface air temperatures are averaged over grid boxes and days for which the daily snowfall is at or above its 99.99th percentile in the control climate (blue solid line; shading shows the interquartile range) and warm climate (red solid line). Mean temperatures are also shown (dashed lines). The blue dashed line deviates from a one-to-one relationship only because of sampling variability. Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.

which is a characteristic snowfall rate at $T = T_m$. The change δs_q also depends inversely on temperature variability as measured by σ^2 , which makes sense given that, for example, temperature variability allows daily temperatures to reach below freezing even if the mean temperature increases to above freezing. Notably, δs_q is independent of the percentile considered, such that the fractional change $\delta s_q/s_q$ is small for sufficiently large s_q . This is the main result from the theory—that the temperature dependencies of precipitation extremes and the rain–snow transition lead to fractional changes in snowfall extremes that are small for sufficiently large snowfall extremes in the control climate. Snowfall extremes respond differently to climate change as compared to precipitation extremes or mean snowfall because snowfall extremes tend to occur at temperatures in a relatively narrow range near the optimal temperature T_m in both the control and warm climates (Fig. 4). As shown schematically in Extended Data Fig. 7, changes in mean temperature do imply changes in the probability of occurrence of temperatures near the optimal temperature for snowfall extremes, but this only results in changes in snowfall extremes that are independent of the percentile considered.

I applied the theory introduced above to the simulations (Methods; Extended Data Fig. 8), and it captures the important features of the response of the snowfall extremes to climate change as a function of climatological monthly temperature (Fig. 2b). (Application of the theory at individual locations is left to future work.) The simulated changes in snowfall extremes asymptote towards the simple theoretical form given by equation (2) as the percentile is increased, and good agreement with the theory is found for the 99.9th and 99.99th percentiles (Extended Data Fig. 9).

Many mountainous regions experience heavy snowfall, but the accuracy of the theory is not as good for regions with surface elevations above 1,000 m (Extended Data Fig. 10), possibly because of variations in the thermodynamic response of orographic precipitation to climate change²³ or the difficulty in simulating orographic snowfall³. Nonetheless, the result that fractional decreases in mean snowfall are greater than those in snowfall extremes seems to hold regardless of elevation in the simulations (Fig. 1 and Extended Data Fig. 10).

Changes in snowfall extremes may still have impacts, and large fractional decreases do occur in the simulations for more moderate extremes and for regions and times of year that are warm enough that there is little snowfall in the control climate (Fig. 2). In addition, changes in the probability of exceeding a fixed high threshold of snowfall (that is, changes in the frequency rather than the intensity of snowfall extremes) may still be substantial because of the exponential tail of precipitation distributions (Extended Data Fig. 7b). Changes in the frequency of snowfall extremes cannot be directly compared with changes in mean snowfall,

but they may be important for impacts that involve a fixed threshold of snowfall. Previous work suggests that the regional climate-change signal of mean snowfall will only emerge after that of temperature, despite large reductions in mean snowfall in many regions^{4,24}. The relatively small fractional changes in snowfall extremes found here suggest that snowfall extremes may not be an early indicator of climate change in many regions; if so, this has corresponding implications for the detection and public perception of climate change.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 January; accepted 27 June 2014.

- Räisänen, J. Warmer climate: less or more snow? *Clim. Dyn.* **30**, 307–319 (2008).
- Brown, R. D. & Mote, P. W. The response of northern hemisphere snow cover to a changing climate. *J. Clim.* **22**, 2124–2145 (2009).
- Kapnick, S. B. & Delworth, T. L. Controls of global snow under a changed climate. *J. Clim.* **26**, 5537–5562 (2013).
- Krasting, J. P., Broccoli, A. J., Dixon, K. & Lanzante, J. Future changes in northern hemisphere snowfall. *J. Clim.* **26**, 7813–7828 (2013).
- Rooney, J. F. Jr. The urban snow hazard in the United States: an appraisal of disruption. *Geogr. Rev.* **57**, 538–559 (1967).
- Kocin, P. J. & Uccellini, L. W. *Northeast Snowstorms* Ch. 1 (American Meteorological Society, 2004).
- Changnon, S. A. & Changnon, D. A spatial and temporal analysis of damaging snowstorms in the United States. *Nat. Hazards* **37**, 373–389 (2006).
- Min, S. K., Zhang, X., Zwiers, F. W. & Hegerl, G. C. Human contribution to more-intense precipitation extremes. *Nature* **470**, 378–381 (2011).
- Westra, S., Alexander, L. V. & Zwiers, F. W. Global increasing trends in annual maximum daily precipitation. *J. Clim.* **26**, 3904–3918 (2013).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119**, 345–357 (2013).
- Trenberth, K. E. Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Clim. Change* **42**, 327–339 (1999).
- O’Gorman, P. A. & Schneider, T. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl Acad. Sci. USA* **106**, 14773–14777 (2009).
- O’Gorman, P. A. & Schneider, T. Scaling of precipitation extremes over a wide range of climates simulated with an idealized GCM. *J. Clim.* **22**, 5676–5685 (2009).
- Kunkel, K. E. et al. Monitoring and understanding trends in extreme storms: state of knowledge. *Bull. Am. Meteorol. Soc.* **94**, 499–514 (2013).
- Zhang, X., Hogg, W. D. & Mekis, É. Spatial and temporal characteristics of heavy precipitation events over Canada. *J. Clim.* **14**, 1923–1936 (2001).
- Kunkel, K. E. et al. Trends in twentieth-century U.S. extreme snowfall seasons. *J. Clim.* **22**, 6204–6216 (2009).
- Diffenbaugh, N. S., Scherer, M. & Ashfaq, M. Response of snow-dependent hydrologic extremes to continued global warming. *Nature Clim. Change* **3**, 379–384 (2013).
- Changnon, S. A., Changnon, D. & Karl, T. R. Temporal and spatial characteristics of snowstorms in the contiguous United States. *J. Appl. Meteorol. Climatol.* **45**, 1141–1155 (2006).
- Kodra, E., Steinhilber, K. & Ganguly, A. R. Persisting cold extremes under 21st-century warming scenarios. *Geophys. Res. Lett.* **38**, L08705 (2011).
- Bourgoin, P. A method to determine precipitation types. *Weath. Forecast.* **15**, 583–592 (2000).
- Dai, A. Temperature and pressure dependence of the rain–snow phase transition over land and ocean. *Geophys. Res. Lett.* **35**, L12802 (2008).
- Feicabrinio, J., Gustafsson, D. & Lundberg, A. Surface-based precipitation phase determination methods in hydrological models. *Hydrol. Res.* **44**, 44–57 (2013).
- Siler, N. & Roe, G. How will orographic precipitation respond to surface warming? An idealized thermodynamic perspective. *Geophys. Res. Lett.* **41**, 2606–2613 (2014).
- Pierce, D. W. & Cayan, D. R. The uneven response of different snow measures to human-induced climate warming. *J. Clim.* **26**, 4148–4167 (2013).

Acknowledgements I thank M. Singh, S. Pfahl, J. Feicabrinio and I. Held for discussions. I am grateful to N. Wood, G. Stephens and the NASA CloudSat project for providing CloudSat snowfall data. I acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and I thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. GPCP 1-degree daily precipitation data were downloaded from <http://www1.ncdc.noaa.gov/pub/data/gpcp/1dd-v1.2/>. NCEP–DOE Reanalysis 2 data were provided by the NOAA/OAR/ESRL PSD at <http://www.esrl.noaa.gov/psd/>. I acknowledge support from NSF grant number AGS-1148594 and NASA ROSES grant number 09-ID509-0049.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to the author (pog@mit.edu).

METHODS

Simulations. The 20 climate models used were BNU-ESM, CanESM2, CMCC-CESM, CMCC-CM, CMCC-CMS, CSIRO-Mk3-6-0, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, HadGEM2-CC, HadGEM2-ES, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC5, MIROC-ESM-CHEM, MIROC-ESM, MPI-ESM-LR, MPI-ESM-MR and MRI-CGCM3. The time period used for HadGEM2-ES for RCP 8.5 was 2081–2099 rather than 2081–2100 because only those years were available in the archive. The first ensemble member was used in all cases.

For Extended Data Fig. 5, a subset of ten models was chosen in which only one model is included from each modelling centre: 1. BNU-ESM, 2. CanESM2, 3. CMCC-CM, 4. CSIRO-Mk3-6-0, 5. GFDL-CM3, 6. HadGEM2-CC, 7. IPSL-CM5A-MR, 8. MIROC5, 9. MPI-ESM-MR, and 10. MRI-CGCM3. These models were selected as either the most recent or the highest resolution in each case.

Calculation of daily snowfall extremes. I calculated snowfall extremes in two ways. In the first method, 20-year return values were calculated from annual maxima using the generalized extreme value (GEV) distribution to allow for relatively-long return periods at each grid box. In the second method, daily snowfall rates were aggregated in bins according to the climatological monthly surface air temperature in the control climate, and high percentiles of snowfall were estimated in each bin; this takes into account the sensitive dependence of snowfall on climatological monthly temperature and allows for a straightforward comparison with theory.

In the first method (Fig. 1 and Extended Data Figs 1 and 2), 20-year return values of daily snowfall were calculated for each model or observational data set, grid box and climate. The 20-year return values were calculated from time series of annual maxima by fitting the GEV distribution using probability-weighted moments²⁵. Probability-weighted moments were used rather than maximum-likelihood estimation because of the relatively short samples, and this approach has been previously used for precipitation extremes²⁶ and to analyse CMIP5 output¹⁰. The goodness of fit was assessed using a Monte Carlo version of the Kolmogorov–Smirnov test²⁶. (A Monte Carlo version of the test was needed because the null hypothesis involves parameters estimated from the time series.) Land grid boxes in the Northern Hemisphere with mean snowfall of greater than 5 cm per year in liquid-water equivalent were considered. The fraction of these grid boxes at which the test was passed at the 10% significance level was found to be close to 10%; the goodness of fit declines if grid boxes with mean snowfall lower than 5 cm per year are included in the analysis. As an additional check, return values were directly estimated as empirical quantiles of the annual maxima time series, and similar results to the GEV estimates were found for a range of quantiles. For the results that are presented as maps, the snowfall statistics were interpolated to a common grid before calculation of multimodel medians. The conclusions are similar if the snowfall extremes are instead measured by the 10-year or 50-year return values (not shown), although the 50-year return values must be viewed with caution given that the underlying time series span roughly 20 years.

In the second method (see Fig. 2), snowfall statistics were analysed as a function of climatological monthly surface air temperature in the control climate. Snowfall extremes were calculated as empirical quantiles of the daily snowfall rates in each temperature bin (without using the GEV distribution in this case). All days, including days with zero snowfall, were included in the analysis. The sample size of snowfall rates in a given temperature bin is of the order of 10^6 , and the 99th, 99.9th and 99.99th percentiles were calculated.

Comparison of simulations with observations. The mean snowfall and snowfall extremes in the simulations are compared with observational estimates in Extended Data Figs 1, 2 and 3. Previous global-scale modelling studies have compared simulated snowfall rates with snowfall rates from reanalysis³ or monthly snowfall rates derived empirically from monthly precipitation rates and monthly surface temperatures⁴. Because observational estimates of daily snowfall are needed and because snowfall from reanalysis may be unreliable³, snowfall rates were estimated here on the basis of observed daily precipitation rates and surface air temperatures and the observed dependence of snowfall fraction on temperature. (Mean snowfall from CloudSat is also discussed below). The precipitation rates are over the period 1997–2012 and were taken from the one-degree daily merged product V1.2 of the Global Precipitation Climatology Project (GPCP 1DD), which includes inputs from infrared, passive microwave, and gauge measurements²⁷. The precipitation rates were first interpolated to a coarser grid with a grid spacing of 2° that is comparable to that of the climate models. Conservative interpolation was used to be consistent with the treatment of precipitation as a flux²⁸. The daily surface air temperatures were taken from the NCEP–DOE reanalysis 2 (NCEP2)²⁹. The dependence of snowfall fraction on temperature was taken from a study of precipitation at Swedish meteorological stations²² (Extended Data Fig. 6) and is given by $\exp[-0.000858(T + 7.5)^{4.12}]$ when the surface air temperature T (in degrees Celsius) is between -4°C and 7°C . All snow was assumed to occur at temperatures below -4°C and all rain at temperatures above 7°C . The snowfall observations are for three-hourly rather than daily accumulations, but this is not expected to affect the results presented particularly

strongly. For example, the good agreement between models and observations shown in Extended Data Fig. 3 is retained if a simple threshold of 1°C is used to determine precipitation type for the GPCP-based observations (that is, assuming all snow below 1°C and all rain above it).

In addition, mean snowfall data from CloudSat³⁰ were used to provide a second and independent comparison with observations (Extended Data Figs 1 and 3). The CloudSat product used (2C-SNOW-PROFILE Release 4) includes vertical profiles of snowfall rate and surface snowfall rate based on reflectivity profiles from the CloudSat Cloud Profiling Radar³¹. The data were available for the period mid-July 2006 to mid-April 2011, which is sufficient to evaluate the mean snowfall rates but is too short to allow for estimation of snowfall extremes.

The overall magnitude and pattern of mean and extreme snowfall are captured by the simulations but with some regional discrepancies (Extended Data Figs 1 and 2). When interpreting the model and observational maps of snowfall, it is important to take into account the area-averaging to a coarse grid and the use of liquid-water equivalent rather than snowfall depth. Snowfall biases in the models may partly relate to temperature biases⁴ and inadequate spatial resolution in regions with high topography³. There are also regional differences in mean snowfall between the two observational estimates (Extended Data Fig. 1), although these differences may relate in part to the different time periods used.

The agreement between the models and the observations is very good when mean and extreme snowfall are analysed as a function of climatological temperature in the control climate (Extended Data Fig. 3). That agreement is better in this case is probably because mean temperature biases are less important when snowfall is analysed as a function of climatological temperature and because variability, circulation biases and random errors are averaged over space and time in each temperature bin. In addition, there is good agreement between the two observational estimates for mean snowfall, except in the lowest temperature bin (Extended Data Fig. 3, bottom panel).

A comparison of the observed snowfall fraction with the snowfall fraction in the simulations (including all surface elevations as in the observations) suggests that the snowfall fraction in the multimodel median is accurate for temperatures below 0°C but declines to zero slightly too quickly for temperatures above 0°C (Extended Data Fig. 6). The discrepancy above 0°C could also result in part from the inexact nature of the comparison between station data and model grid boxes and from the difficulty of apportioning mixed snow and rain in observations. This discrepancy does not affect the optimal temperature T_m in the theory of snowfall extremes because $T_m < 0^\circ\text{C}$. Note that the rain–snow transition does not occur precisely at a surface temperature of 0°C because frozen precipitation does not immediately melt as it falls past the melting level and because of temperature variability within the accumulation period used.

Derivation of theory for snowfall extremes. The following assumptions are made in the derivation, as discussed in the main text. The daily snowfall rate s is related to the daily precipitation rate p and daily surface air temperature T according to $s = f(T)p$, where $f(T)$ is the fraction of precipitation that falls as snow at a given temperature T . The daily surface air temperature T is assumed to be normally distributed with mean \bar{T} and standard deviation σ . The precipitation rate p has a simple dependence on T according to $p = e^{\beta T} \hat{p}$. This exponential dependence on temperature is motivated by the thermodynamic scaling of precipitation extremes under climate change¹² and the observed covariability of daily precipitation extremes with surface temperature³². The normalized precipitation rate \hat{p} is assumed to follow a gamma distribution on wet days³³, such that its probability density function P is given by:

$$P(\hat{p}) = (1 - w)\delta(\hat{p}) + \frac{w\gamma^k}{\Gamma(k)} \hat{p}^{k-1} e^{-\gamma\hat{p}}$$

where δ is the delta function, Γ is the gamma function, w is the fraction of wet days, $1/\gamma$ is the scale parameter and k is the shape parameter. (When applying the theory to the simulations, wet days are defined as days with precipitation greater than 0.1 mm per day rather than precipitation greater than zero, as described here.) The temperature T and the normalized precipitation rate \hat{p} are assumed to be independent.

With these assumptions, the q th percentile of snowfall s_q is exceeded if the following inequality is satisfied:

$$\hat{p} e^{\beta T} f(T) > s_q$$

which requires that $\hat{p} > h(T)s_q$, where $h(T) = e^{-\beta T} f(T)^{-1}$. Assuming s_q is non-zero, the probability that s_q is exceeded may be written as:

$$1 - \frac{q}{100} = \int_{-\infty}^{\infty} dT \int_{h(T)s_q}^{\infty} d\hat{p} \frac{w\gamma^k}{\Gamma(k)} \hat{p}^{k-1} e^{-\gamma\hat{p}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T-\bar{T})^2}{2\sigma^2}} \quad (3)$$

Asymptotic methods are next used to evaluate the double integral in equation (3) in the extreme snowfall limit of large s_q . The integral in \hat{p} is first evaluated using

a standard asymptotic expression for the incomplete gamma function³⁴:

$$\int_z^\infty dt t^{k-1} e^{-t} = z^{k-1} e^{-z} [1 + O(z^{-1})] \quad (4)$$

in the limit of large and positive z . Making the identifications

$$t = \gamma \hat{p}$$

$$z = \gamma h s_q$$

in equation (4) gives:

$$\int_{\gamma h s_q}^\infty d(\gamma \hat{p}) (\gamma \hat{p})^{k-1} e^{-\gamma \hat{p}} = (\gamma h s_q)^{k-1} e^{-\gamma h s_q} \quad (5)$$

which is valid asymptotically for large s_q . (The tilde symbol to denote 'is asymptotic to' is not used here, to avoid confusion with its common use to denote scaling behaviour.) Note that $\gamma > 0$ and $h(T) > 0$. Substituting equation (5) into equation (3) gives:

$$1 - \frac{q}{100} = \frac{(\gamma s_q)^{k-1} w}{\Gamma(k) \sqrt{2\pi\sigma}} \int_{-\infty}^\infty dT h(T)^{k-1} e^{-\gamma h(T) s_q - \frac{(T-\bar{T})^2}{2\sigma^2}} \quad (6)$$

For large s_q , the integral in temperature is dominated by the contribution close to $T = T_m$ at which $h(T)$ reaches a minimum, which corresponds physically to snowfall extremes occurring near the optimal temperature T_m (found to be roughly -2°C). The integral may be evaluated asymptotically using Laplace's method³⁵, and the general result used here is:

$$\int_{-\infty}^\infty dt g(t) e^{x\phi(t)} = \sqrt{\frac{2\pi}{-x\phi''(c)}} g(c) e^{x\phi(c)} [1 + O(x^{-1})] \quad (7)$$

as $x \rightarrow \infty$, where the function ϕ reaches a maximum at $t = c$, and the first and second derivatives of ϕ are denoted ϕ' and ϕ'' , respectively. Here

$$x = \gamma s_q$$

$$t = T$$

$$c = T_m$$

$$g(t) = h(T)^{k-1} e^{-\frac{(T-\bar{T})^2}{2\sigma^2}}$$

$$\phi(t) = -h(T)$$

and ϕ reaches a maximum when h reaches a minimum. These substitutions are used in equation (7) to give:

$$\int_{-\infty}^\infty dT h(T)^{k-1} e^{-\gamma h(T) s_q - \frac{(T-\bar{T})^2}{2\sigma^2}} = \sqrt{\frac{2\pi}{\gamma s_q h_m''}} h_m^{k-1} e^{-\gamma s_q h_m - \frac{(T_m - \bar{T})^2}{2\sigma^2}}$$

which is valid asymptotically for large s_q , and where the subscript 'm' refers to a quantity evaluated at $T = T_m$. Substituting this into equation (6) yields equation (1) in the main text. Equation (1) can always be solved for s_q if $k < \frac{3}{2}$, as is generally the case in the simulations (Extended Data Fig. 8).

Derivation of simple expression for changes in snowfall extremes. The change in s_q may be calculated by evaluating s_q from equation (1) in each climate and taking the difference. Alternatively, a simple expression is derived here for the change in s_q , assuming that all parameters other than the mean temperature \bar{T} remain constant. The changes in s_q and \bar{T} between the control and warm climate are denoted δs_q and $\delta \bar{T}$, respectively. Taking the ratio of the left-hand side of equation (1) in the warm and control climates and equating it to the same ratio for the right-hand side yields:

$$\frac{(s_q + \delta s_q)^{\frac{3}{2}-k}}{s_q^{\frac{3}{2}-k}} e^{\gamma h_m \delta s_q} = e^{-\frac{1}{2\sigma^2}[(\bar{T} + \delta \bar{T} - T_m)^2 - (\bar{T} - T_m)^2]}$$

Taking the logarithm and rearranging terms gives:

$$\frac{\delta s_q}{s_q} = -\frac{\delta \bar{T}}{\sigma^2 \gamma h_m s_q} \left(\bar{T} + \frac{\delta \bar{T}}{2} - T_m \right) + \frac{k - \frac{3}{2}}{\gamma h_m s_q} \log \left(1 + \frac{\delta s_q}{s_q} \right) \quad (8)$$

Since the limit of $s_q \rightarrow \infty$ is being taken, equation (8) implies that $\delta s_q/s_q \rightarrow 0$. The alternative limits $\delta s_q \rightarrow -s_q$ or $\delta s_q/s_q \rightarrow \infty$ in which the logarithm on the right-hand

side of equation (8) becomes large in magnitude are inconsistent with equation (8) because $k < \frac{3}{2}$. Because $\delta s_q/s_q \rightarrow 0$, the second term on the right-hand side of equation (8) may then be neglected, and equation (2) in the main text is obtained.

According to equation (2), the change in snowfall extremes is independent of q , w , k and h_m'' . If it is found that $\delta s_q < -s_q$ when applying equation (2), then the starting point given by equation (3) is invalid because it assumes $s_q > 0$, and we must instead set $\delta s_q = -s_q$. Note that unlike equation (2), equation (1) has the accidental advantage of always implying non-negative snowfall rates even when the assumptions made in its derivation are not accurate.

Application of the theory to the simulations. The snowfall fraction $f(T)$ is needed to calculate $h(T)$ and the optimal temperature T_m . It was calculated for each model and climate by binning the daily precipitation and snowfall rates in surface air temperature bins of 0.25°C over land in the Northern Hemisphere and below or above 1,000 m elevation as required (Fig. 3). Because the second derivative of $h(T)$ was needed, the diagnosed $f(T)$ was smoothed using a Gaussian filter with standard deviation 0.5°C before calculation of $h(T)$. The multimodel medians of T_m and $f(T_m)$ are -2.3°C and 0.89, respectively, in both the control climate and warm climate, for the default case of surface elevations below 1,000 m. The functional fit to the snowfall fraction from observations²² discussed earlier yields similar values of $T_m = -2.3^\circ\text{C}$ and $f(T_m) = 0.93$.

The parameter describing the thermodynamic dependence of precipitation extremes was set to $\beta = 0.06^\circ\text{C}^{-1}$, following previous work¹². The other parameters in the theory were evaluated for each control-climate temperature bin using the temperatures and precipitation rates aggregated within the bin. Wet days were defined to occur when precipitation is at or above 0.1 mm per day, and the gamma distribution was fitted to wet-day values of \hat{p} using the method of moments to estimate γ and k (Extended Data Fig. 8).

The theory tends to underestimate the absolute magnitudes of the snowfall extremes for the 99.99th percentile (Extended Data Fig. 4), although the fractional changes between climates are still accurate (Fig. 2b). The underestimate of the absolute magnitudes of the 99.99th percentiles results primarily from inaccuracies in the fit of the gamma distribution to the distribution of \hat{p} . The method of moments was used to fit the gamma distribution because it was found to give a better fit than maximum-likelihood estimation for the moderate and extreme parts of the \hat{p} distribution. One potential change to the theory would be to fit alternative distributions³⁶ needed to calculate the snowfall extremes and thus would not lead to a simple result. In the section 'Alternative form of theory using Weibull distribution' below, I show that the theory may still be evaluated asymptotically when the Weibull distribution is used instead of the gamma distribution. The conclusions are similar, with the primary difference being that greater deviations from an exponential tail are possible than with the gamma distribution, and these deviations can lead to a weak dependence of the changes in snowfall extremes on the percentile considered.

The theory also assumes that \hat{p} (a proxy for upward motion) and temperature are independent, but upward motion and precipitation are generally less likely to occur on anomalously cold days³⁷, and the accuracy of the theory could be improved by accounting for this relationship. This refinement to the theory is not attempted here because of the additional complexity and assumptions needed and because the current form of the theory adequately captures the main features of the response of daily snowfall extremes to climate change.

Alternative form of theory using Weibull distribution. The theory is also tractable if the normalized precipitation rate \hat{p} is assumed to follow a Weibull distribution on wet days instead of a gamma distribution. The probability density function P for \hat{p} is then given by:

$$P(\hat{p}) = (1-w)\delta(\hat{p}) + w l \alpha (x\hat{p})^{l-1} e^{-(x\hat{p})^l}$$

where δ is the delta function, w is the fraction of wet days, $1/\alpha$ is the scale parameter, and l is the shape parameter. In calculating the q th percentile of snowfall, the integral in \hat{p} is exact, and the integral in T is performed using Laplace's method as before. The result is:

$$(\alpha s_q h_m)^{1/2} e^{(\alpha s_q h_m)^l} = \frac{w}{\sigma(1-\frac{q}{100})} \sqrt{\frac{h_m^l}{(h^l)_m}} e^{-\frac{(T-\bar{T})^2}{2\sigma^2}}$$

The simple expression for the change in s_q , corresponding to equation (2) when the gamma distribution is used, is given by:

$$\delta s_q = -\frac{\delta \bar{T}}{\sigma^2 l (\alpha h_m)^l (s_q)^{l-1}} \left(\bar{T} + \frac{\delta \bar{T}}{2} - T_m \right) \quad (9)$$

The parameters in the Weibull distribution were estimated using maximum-likelihood estimation, and the results for the changes in snowfall extremes were

found to be similar to the results from the theory using the gamma distribution (not shown). According to equation (9), the change in snowfall extremes δs_q depends on $(s_q)^{l-1}$ and therefore is no longer completely independent of the percentile to the extent that l differs from 1. However, this dependence was found to be weak, and typical values of l in the simulations are in the range 0.7–1.1. Importantly, it is still the case that the fractional change $\delta s_q/s_q$ is small for sufficiently large s_q because $l > 0$.

Role of circulation changes and robustness of results. In the theory, γ and k are the parameters that are most strongly tied to dynamics and updraft strength. These parameters do change to some extent as the climate warms (Extended Data Fig. 8), but they do not change sufficiently to alter the large contrast between the changes in mean and extreme snowfall, and similar results are found whether snowfall extremes are estimated from the full theory given by equation (1) or if the simple estimate given by equation (2) is used that assumes parameters such as γ and k are fixed (Fig. 2b). The ratios from equation (2) are calculated as $1 + \delta s_q/s_q$, where all parameters other than the temperature change are evaluated from the control climate.

Much of the uncertainty in changes in upward velocities in climate-model simulations is thought to relate to parameterized moist convection^{38,39}, which is more important for warm-season or tropical precipitation, even if convection may enhance snowfall locally in a given storm. Consistent with this interpretation, extratropical precipitation extremes are generally found to respond to climate change in a robust manner, unlike tropical precipitation extremes^{12,39}. Inaccuracy in simulating Arctic sea-ice loss could affect the warming pattern and circulation, but this would not be expected to alter the contrast between the responses of mean and extreme daily snowfall substantially, and similar results are found here for the subset of models that have previously been identified⁴⁰ as performing well when simulating Arctic sea ice (not shown).

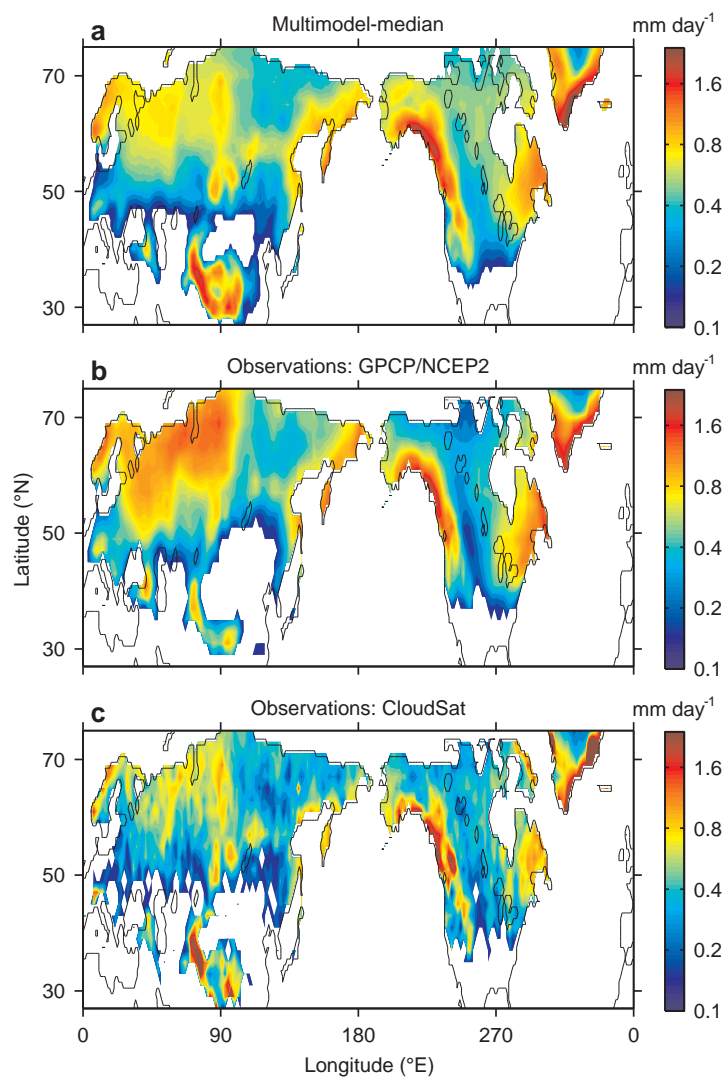
Extended Data Fig. 5 illustrates the robustness of the greater declines in mean snowfall as compared to snowfall extremes. To increase the extent to which the models are independent, a subset of ten models with only one model from each climate centre was analysed (see the ‘Simulations’ section for the list of models). Extended Data Fig. 5a shows that there are widespread regions in which snowfall extremes (as measured by the 20-year return period) fractionally decrease by less than mean snowfall (or increase) in all ten of the models considered. Extended Data Fig. 5b shows that for each of the models separately the fractional decrease in snowfall extremes is robustly less than that in mean snowfall for the -2.5°C control-climate temperature bin.

Heuristic argument for changes in snowfall extremes. The simple estimate given by equation (2) may also be obtained using a heuristic argument based on the property that snowfall extremes tend to occur at temperatures close to T_m in both the control and warm climates (Fig. 4). Consider the case, illustrated in Extended Data Fig. 7, in which the mean temperature is above T_m in the control climate. The joint probability density function (PDF) of temperature T and normalized precipitation \hat{p} is the product of a Gaussian distribution in temperature and a gamma distribution in \hat{p} . An increase in mean temperature reduces the joint PDF in the preferred temperature range for extreme snowfall near T_m (Extended Data Fig. 7a), with the result that high percentiles of \hat{p} and snowfall must also decrease (Extended Data Fig. 7b). The integral of the joint PDF over $\hat{p} > s_q h_m$ at $T = T_m$ must remain approximately the same in each climate because the percentile considered is unchanged. At $T = T_m$, the joint PDF has an exponential dependence on $-(T_m - \bar{T})^2 / (2\sigma^2) - \gamma \hat{p}$, and considering only the exponential part for simplicity, we find that $-\delta \left[(T_m - \bar{T})^2 / (2\sigma^2) \right] - \gamma \delta s_q h_m = 0$. In the limit of a small change in mean temperature, we find that $\delta s_q = \delta \bar{T} (T_m - \bar{T}) / (\sigma^2 \gamma h_m)$, which is consistent with equation (2). So the increase in mean temperature reduces the snowfall extremes in this

case, but by an amount that is independent of the percentile considered, such that the change is a small fraction of the snowfall extreme in the control climate for sufficiently high percentiles.

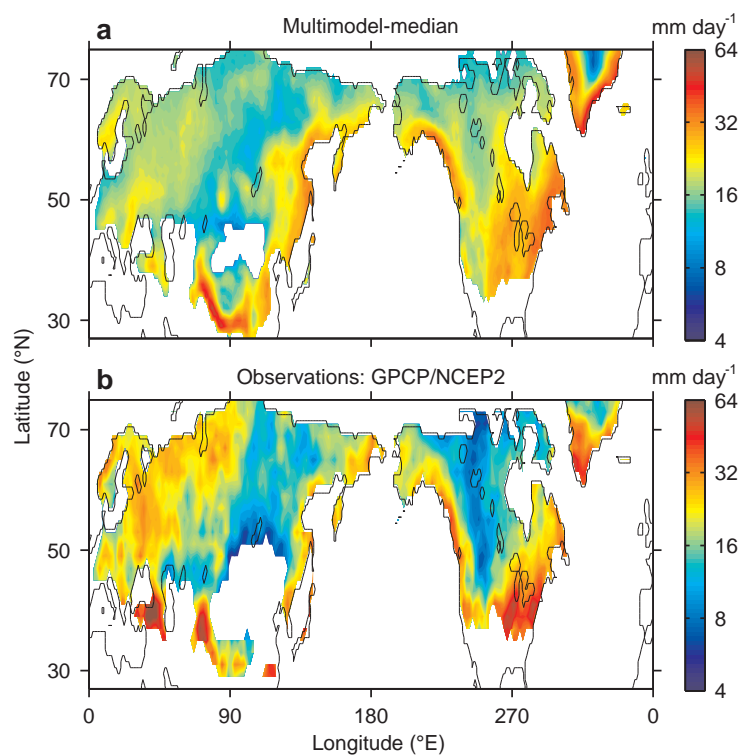
Snowfall depth versus liquid-water equivalent. Snowfall is expressed in liquid-water equivalent in the simulations, but snowfall depth is often measured in observations⁴¹. Snowfall depth depends on snow density in addition to the liquid-water equivalent, and snow density depends on temperature as well as other factors. The theory of snowfall extremes described above may be easily modified to apply to snowfall depth by assuming a functional dependence of snow density on temperature and including this dependence in the expression relating snowfall and precipitation rates. The snowfall extremes measured in snowfall depth would then be associated with a lower optimal temperature T_m than those measured in liquid-water equivalent (for example, using equations (1) and (2) of ref. 42 for the density of snow together with the observed snowfall fraction curve²² yields $T_m = -4.3^\circ\text{C}$), but the basic features of the contrast between the responses of mean snowfall and snowfall extremes remain the same.

25. Hosking, J. R. M., Wallis, J. R. & Wood, E. F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261 (1985).
26. Kharin, V. V. & Zwiers, F. W. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.* **13**, 3760–3788 (2000).
27. Huffman, G. J. *et al.* Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.* **2**, 36–50 (2001).
28. Chen, C. T. & Knutson, T. On the verification and comparison of extreme rainfall indices from climate models. *J. Clim.* **21**, 1605–1621 (2008).
29. Kanamitsu, M. *et al.* NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**, 1631–1643 (2002).
30. Stephens, G. L. *et al.* CloudSat mission: performance and early science after the first year of operation. *J. Geophys. Res.* **113**, D00A18 (2008).
31. Wood, N. Level 2c snow profile process description and interface control document, version 0. <http://www.cloudsat.cira.colostate.edu/datalCDlist.php?go=list&path=/2C-SNOW-PROFILE> (2013).
32. Lenderink, G. & van Meijgaard, E. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geosci.* **1**, 511–514 (2008).
33. Groisman, P. Y. *et al.* Changes in the probability of heavy precipitation: Important indicators of climatic change. *Clim. Change* **42**, 243–283 (1999).
34. Abramowitz, M. & Stegun, I. A. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* Vol. 55, section 6.5.32 (Dover, 1964).
35. Bender, C. M. & Orszag, S. A. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* Vol. 1, Ch. 6 (Springer, 1999).
36. Furrer, E. M. & Katz, R. W. Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resour. Res.* **44**, W12439 (2008).
37. de Vries, H., Haarsma, R. J. & Hazeleger, W. On the future reduction of snowfall in western and central Europe. *Clim. Dyn.* **41**, 2319–2330 (2013).
38. Wilcox, E. M. & Donner, L. J. The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. *J. Clim.* **20**, 53–69 (2007).
39. O’Gorman, P. A. Sensitivity of tropical precipitation extremes to climate change. *Nature Geosci.* **5**, 697–700 (2012).
40. Massonnet, F. *et al.* Constraining projections of summer Arctic sea ice. *Cryosphere* **6**, 1383–1394 (2012).
41. Paulhus, J. L. H. Record snowfall of April 14–15, 1921, at Silver Lake, Colorado. *Mon. Weath. Rev.* **81**, 38–40 (1953).
42. Brown, R. D., Brasnett, B. & Robinson, D. Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere–Ocean* **41**, 1–14 (2003).



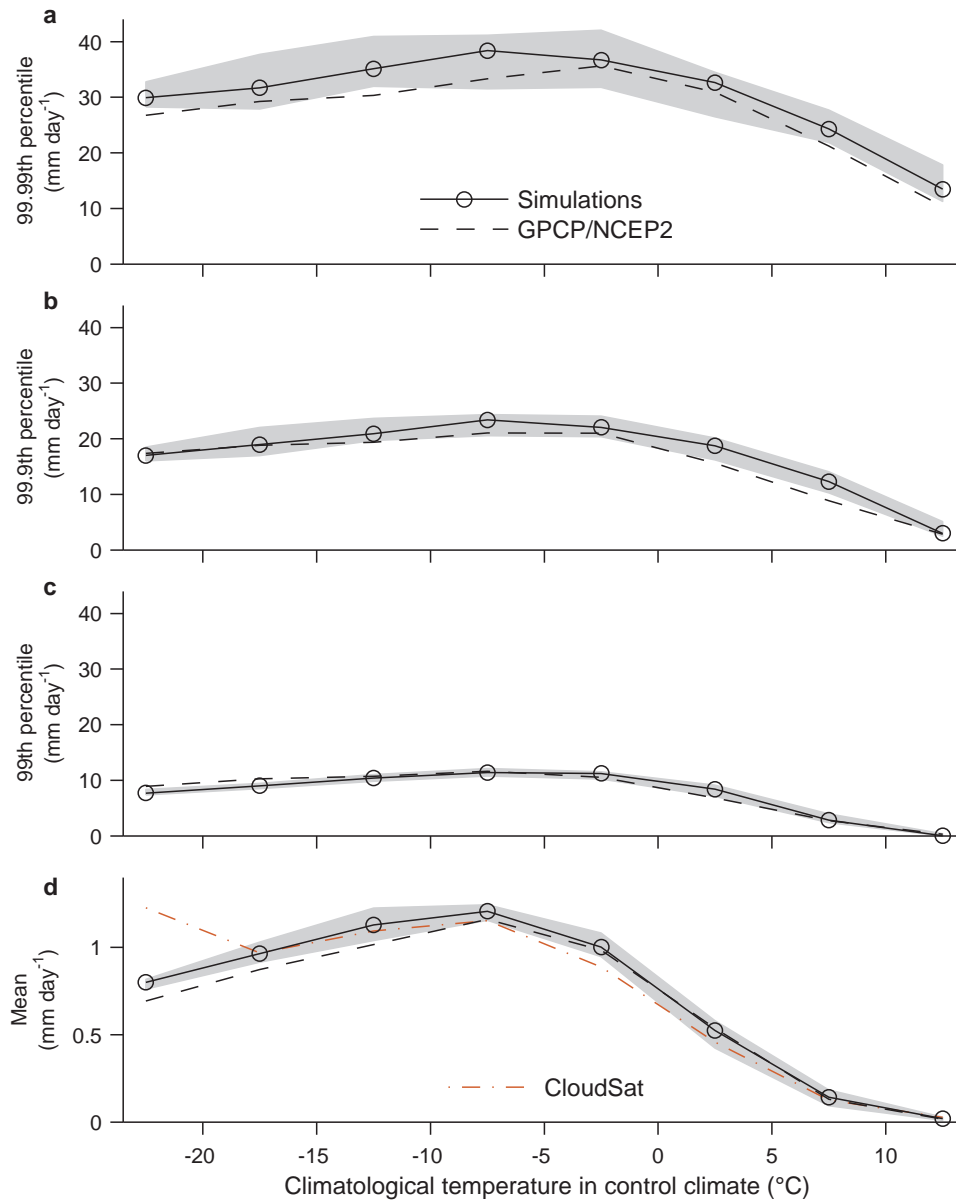
Extended Data Figure 1 | Mean snowfall in simulations and observations.
a, The control climate in the multimodel median. **b**, **c**, Observational estimates

from GPCP/NCEP2 (**b**) and CloudSat (**c**). In each case, results are only shown where mean snowfall exceeds 5 cm per year.



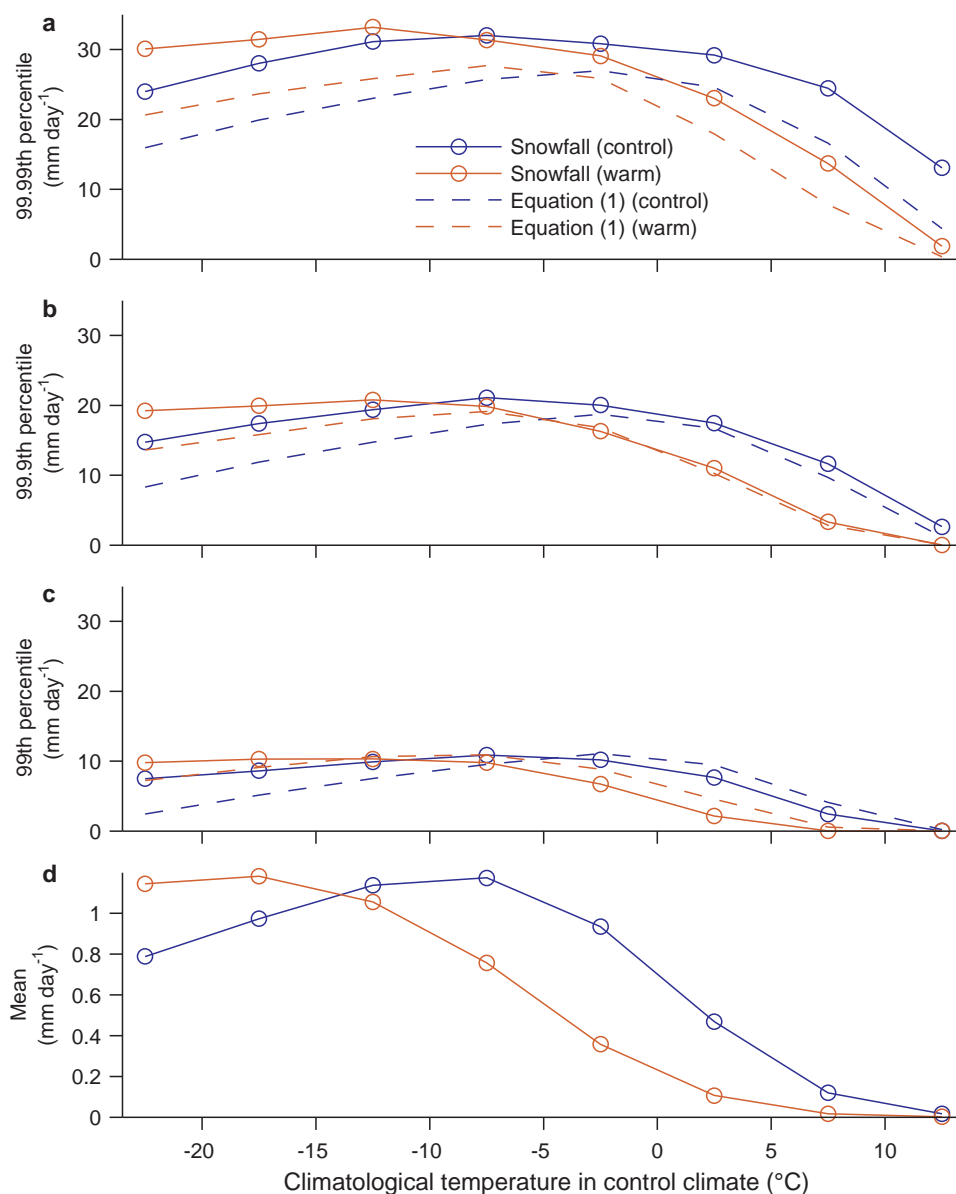
Extended Data Figure 2 | Daily snowfall extremes in simulations and observations. **a**, The control climate in the multimodel median. **b**, Observational estimate from GPCP/NCEP2. The snowfall extremes shown

are the 20-year return values estimated using a fit of the generalized extreme value distribution to the annual-maximum time series. In each case, results are only shown where mean snowfall exceeds 5 cm per year.



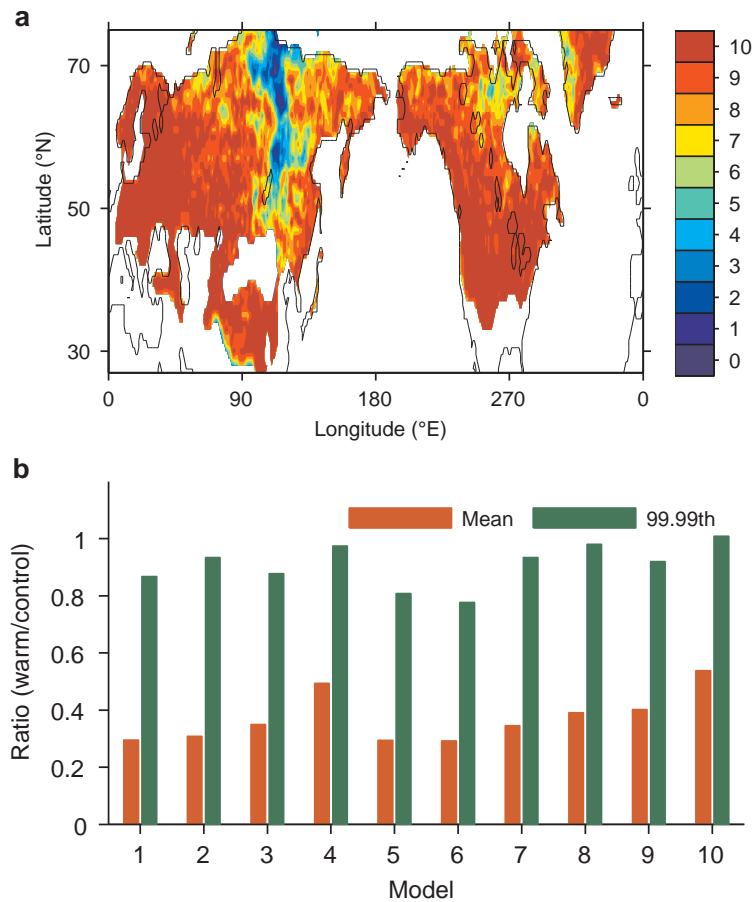
Extended Data Figure 3 | Mean and extreme snowfall as a function of climatological monthly surface air temperature in simulations and observations. **a–d,** The 99.99th (a), 99.9th (b), and 99th (c) percentiles of daily snowfall and mean snowfall (d) are shown for the control climate in the multimodel median (black solid line with circles; shading shows the

interquartile range) and as estimated from GPCP/NCEP2 (black dashed line). CloudSat mean snowfall (red dashed-dotted line) is also shown in **d**. For the observational curves, NCEP2 monthly temperatures were used to define the climatological monthly surface air temperature bins. Only land grid boxes in the Northern Hemisphere (but all surface elevations) are included.



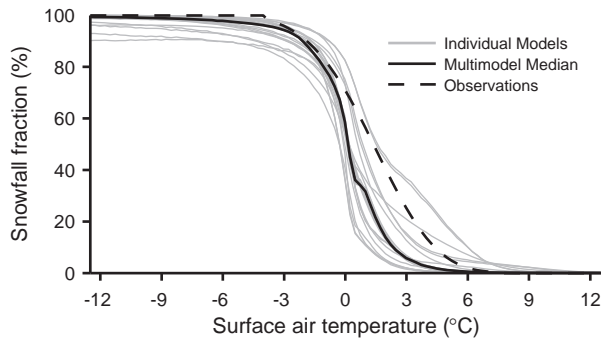
Extended Data Figure 4 | Mean and extreme snowfall in different climates as a function of climatological monthly surface air temperature. a–d, The multimodel-median 99.99th (a), 99.9th (b), and 99th (c) percentiles of daily snowfall and mean snowfall (d) are shown in the control climate (blue line with circles) and warm climate (red line with circles). The snowfall statistics

shift left with warming (to some extent) because of the important influence of temperature on snowfall. Also shown are theoretical estimates given by equation (1) for high percentiles of snowfall in the control climate (blue dashed line) and the warm climate (red dashed line). Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.

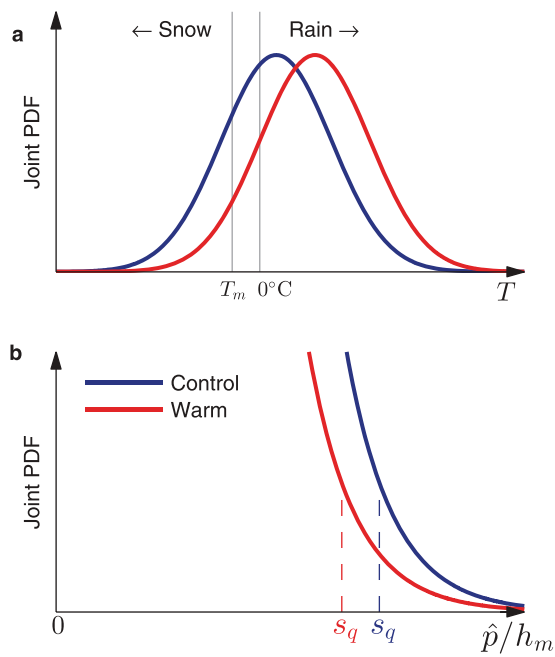


Extended Data Figure 5 | Robustness of greater declines in mean snowfall compared with snowfall extremes in ten models from different centres (see Methods). **a**, Number of models out of ten in which the fractional decrease in the 20-year return value is less than that for mean snowfall or the 20-year return value increases. **b**, Ratios of mean snowfall (red) and the 99.99th percentile of daily snowfall (green) for the warm climate compared to the

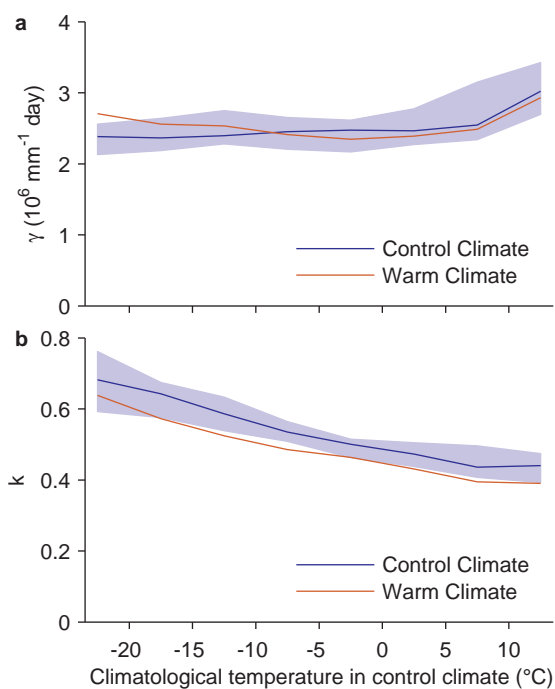
control climate and the -2.5°C control-climate temperature bin. In **a**, only land grid boxes with mean snowfall greater than 5 cm per year in the control climate in the multimodel median are shown. In **b**, only Northern-Hemisphere land grid boxes with surface elevation below 1,000 m are included, and the models are identified by number in the 'Simulations' section of the Methods.



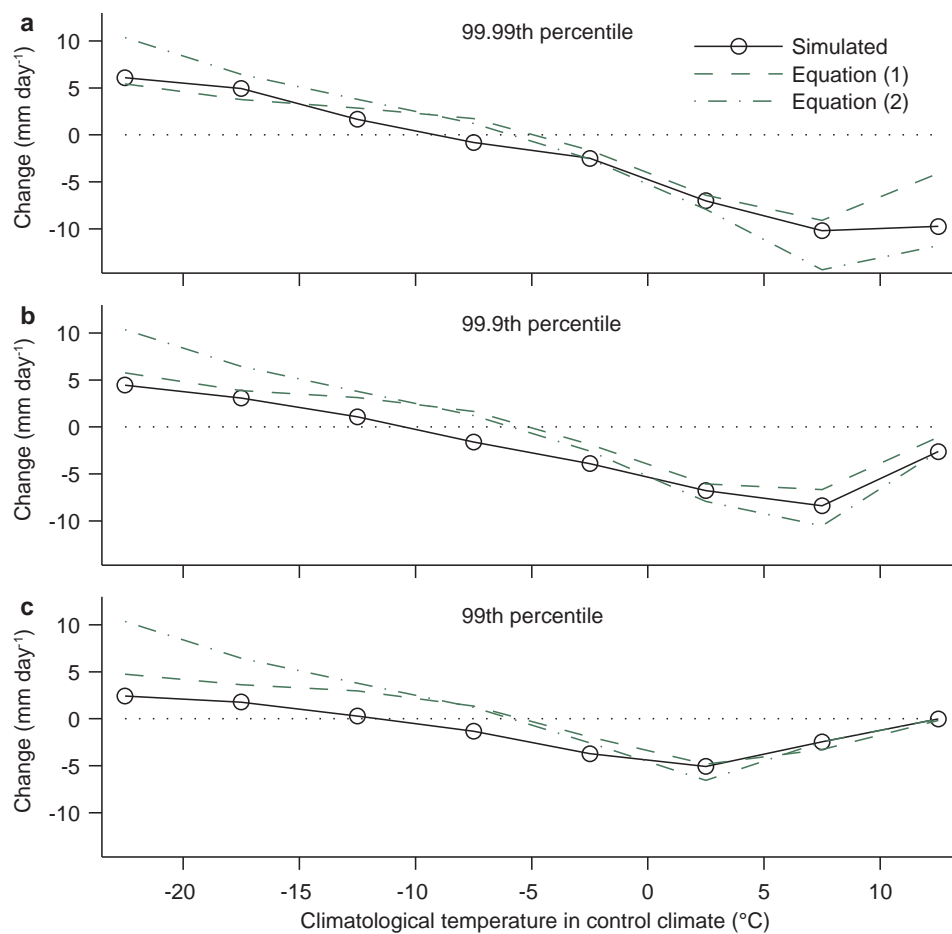
Extended Data Figure 6 | Snowfall fraction as a function of surface air temperature in simulations and observations. The snowfall fraction is shown for the control climate in individual models (grey lines) and the multimodel median (solid black line). A functional fit to observations is shown for comparison (black dashed line). The snowfall fraction for models is calculated as the ratio of mean snowfall to mean precipitation in daily temperature bins of width 0.25°C , as in Fig. 3 but with all surface elevations included. The functional fit to the observed snowfall fraction is for three-hourly observations from Swedish meteorological stations²².



Extended Data Figure 7 | Schematic illustrating the effect of climate warming on the joint PDF of temperature T and normalized precipitation rate \hat{p} , and the resulting change in a high snowfall percentile s_q . **a, b,** The joint PDF as a function of T at a fixed \hat{p} (**a**), and as a function of snowfall rate \hat{p}/h_m at $T = T_m$ close to which snowfall extremes tend to occur (**b**). The joint PDF is shown for the control (blue) and warm (red) climates. Mean snowfall and the probability of snowfall can be inferred to decrease markedly with warming from **a**, while in **b** the area under the joint PDF to the right of s_q is the same in each climate, and s_q experiences a relatively small fractional decrease with warming.

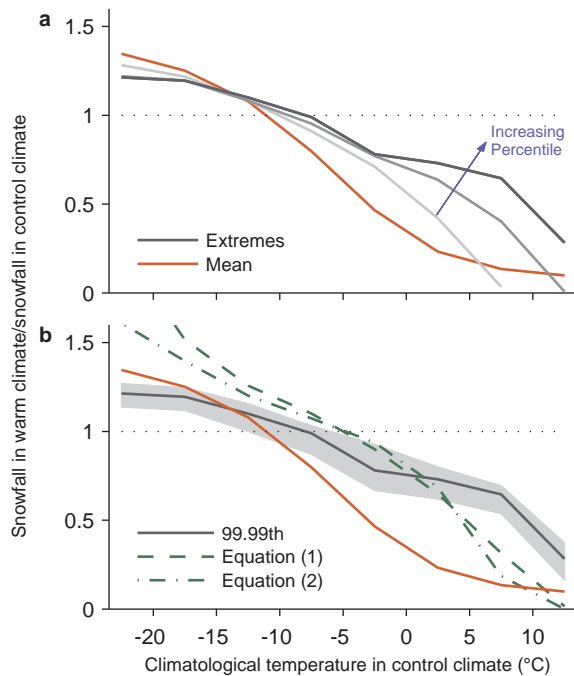


Extended Data Figure 8 | Parameters in the theory as a function of climatological monthly surface air in the control climate. a, b, Shown are the multimodel-medians of the rate parameter γ (a) and shape parameter k (b) in the control climate (blue line; shading shows the interquartile range) and warm climate (red line). Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.



Extended Data Figure 9 | Multimodel-median changes in snowfall extremes between the control and warm climates as a function of climatological monthly surface air temperature in the control climate. a–c, 99.99th (a), 99.9th (b) and 99th (c) percentiles of daily snowfall for the simulations (black line with circles), theory estimate from equation (1) (green dashed line),

and simple theory estimate from equation (2) (green dashed-dotted line). The simple theory estimate is not independent of percentile for high climatological temperatures because it is constrained to not imply a negative snowfall rate in the warm climate. Only land grid boxes in the Northern Hemisphere with surface elevation below 1,000 m are included.



Extended Data Figure 10 | Ratios of snowfall for land grid boxes in the Northern Hemisphere with elevations at or above 1,000 m. **a, b,** Ratios are shown for the warm climate compared with the control climate as a function of climatological monthly surface air temperature in the control climate. Multimodel-median ratios of mean snowfall (red line) are shown in both panels. **a,** Multimodel-median ratios of the 99th, 99.9th and 99.99th percentiles of daily snowfall in increasing order from light to dark grey. **b,** Multimodel-median ratio of the 99.99th percentile of daily snowfall (grey line; shading shows the interquartile range), and the same ratio according to the theory estimate from equation (1) (green dashed line) and the simple theory estimate from equation (2) (green dashed-dotted line).

A primitive fish from the Cambrian of North America

Simon Conway Morris¹ & Jean-Bernard Caron^{2,3}

Knowledge of the early evolution of fish largely depends on soft-bodied material from the Lower (Series 2) Cambrian period of South China^{1,2}. Owing to the rarity of some of these forms and a general lack of comparative material from other deposits, interpretations of various features remain controversial^{3,4}, as do their wider relationships amongst post-Cambrian early un-skeletonized jawless vertebrates. Here we redescribe *Metaspriggina*⁵ on the basis of new material from the Burgess Shale and exceptionally preserved material collected near Marble Canyon, British Columbia⁶, and three other Cambrian Burgess Shale-type deposits from Laurentia. This primitive fish displays unambiguous vertebrate features: a notochord, a pair of prominent camera-type eyes, paired nasal sacs, possible cranium and arcualia, W-shaped myomeres, and a post-anal tail. A striking feature is the branchial area with an array of bipartite bars. Apart from the anterior-most bar, which appears to be slightly thicker, each is associated with externally located gills, possibly housed in pouches. Phylogenetic analysis places *Metaspriggina* as a basal vertebrate, apparently close to the Chengjiang taxa *Haikouichthys*^{1–4} and *Mylokunmingia*¹, demonstrating also that this primitive group of fish was cosmopolitan during Lower–Middle Cambrian times (Series 2–3). However, the arrangement of the branchial region in *Metaspriggina* has wider implications for reconstructing the morphology of the primitive vertebrate. Each bipartite bar is identified as being respectively equivalent to an epibranchial and ceratobranchial. This configuration suggests that a bipartite arrangement is primitive and reinforces the view that the branchial basket of lampreys⁷ is probably derived. Other features of *Metaspriggina*, including the external position of the gills and possible absence of a gill opposite the more robust anterior-most bar, are characteristic of gnathostomes⁸ and so may be primitive within vertebrates.

Cambrian chordates^{1–5} are effectively restricted to the Burgess Shale^{5,9} and Chengjiang^{1–4} Lagerstätten, and despite soft-part preservation key structures may be difficult to resolve^{3,9,10}. In addition, differential decay of various parts of the body may skew interpretation of character states¹¹ and potentially compromise phylogenetic analysis.

Here we redescribe the poorly known Burgess Shale chordate *Metaspriggina walcotti*⁵ (Extended Data Fig. 1), on the basis of approximately 100 new specimens collected in the Canadian Rockies from several Cambrian Burgess Shale-type deposits. Localities include the Burgess Shale (Walcott Quarry) in Yoho National Park (Fig. 1g and Extended Data Fig. 2), and three sites in Kootenay National Park: Haiduk Peak (Extended Data Fig. 3)¹² and, most importantly, near Marble Canyon⁶ (Fig. 1a–f, h–m and Extended Data Figs 4, 5). In addition, *Emmonsaspis (partim)* from the Parker Slate Formation¹³ in Vermont, hitherto tentatively identified as a frond-like fossil¹⁴, and the Kinzers Formation in Pennsylvania (R. Thomas, personal communication), are reinterpreted as *Metaspriggina* spp. (Extended Data Fig. 6). Despite variations in age (Extended Data Fig. 7 and Supplementary Table 1) and palaeoenvironmental settings between these occurrences (and by implication taphonomic histories) the similarities of anatomy allow reliable reconstructions.

Descriptions are largely based on the more complete ‘Marble Canyon’⁶ fossils (see also Fig. 2). Like other Burgess Shale-type material, fossils are preserved as carbonaceous compressions and aluminosilicate minerals

(Extended Data Fig. 4f). Incompleteness precludes a precise estimate of size range, but the most complete specimens (Fig. 1a, b) are about 60 mm in length and 8–13 mm in height. Laterally the body is fusiform, widest near the middle, tapering to a fine point posteriorly (Fig. 1a, b and Extended Data Fig. 4a), whereas in dorsal view the anterior termination is rounded (Fig. 1d and Extended Data Fig. 4c–e). The animal was compressed laterally, as is evident from occasional folding of the body as well as specimens in dorso-ventral orientation being conspicuously narrower (Fig. 1a and Extended Data Fig. 5a). Along the anterior ventral margin there was a keel-like structure (Fig. 1b, g, i, k, l), but no fins have been recognized. In the much more abundant specimens of *Haikouichthys*^{1,3,4} fins are seldom obvious, suggesting that their absence in *Metaspriggina* might be taphonomic.

The diagnostic myomeres are invariably present, but are sometimes highly disorganized especially towards the anterior (Fig. 1b, d, i and Extended Data Fig. 1g), suggesting variable levels of decay¹¹. Undisrupted specimens show open chevrons with the main apices directly anteriorly (Fig. 1h, k and Extended Data Fig. 5c). Dorsally, however, some specimens show a smaller additional fold with the blunt apex directed posteriorly (Fig. 1k), while in one specimen (Fig. 1h) another posterior inflection lies towards the ventral side. Overall, therefore, the myomeres have a W-shaped configuration. As the body narrows posteriorly the principal myotomal apices become much more acute (approximately 60°), and the subsidiary dorsal inflection is probably lost (Extended Data Figs 1g, i, 4a). The myomeres, totalling at least 40, are considerably more acute than in *Pikaia*⁹ and, in contrast to this chordate, *Metaspriggina* was evidently an effective swimmer¹⁵. Parker Slate specimens differ in being less slender and having myomeres with a more angular closure (Extended Data Fig. 6a–d). In dorso-ventral specimens an elongate strand (0.25 mm thick) is identified as the notochord (Fig. 1a–d and Extended Data Fig. 1a–d, f). In laterally orientated material the notochord lies on the midline, opposite the zone of myotomal closure (Fig. 1h). Occasional narrower strands in the anterior region probably represent parts of the vascular system (Fig. 1d).

The head is small and slightly bilobed, with smooth margins and possibly a central notch (Fig. 1d and Extended Data Fig. 4c–e). It bears two prominent eyes (Fig. 1a–d, f, g, i, j, l, m and Extended Data Figs 3, 4b–e, 5a–d, 6d). These appear to have been originally circular (maximum diameter approximately 1.3 mm), but elliptical shapes (Fig. 1d) may reflect slightly oblique burial. Typically the eye is preserved as a reflective film, but a well-defined circular area (Fig. 1f) (approximately 0.4 mm) appears to be the lens, suggesting a camera-like arrangement. In several specimens (Fig. 1c, i, j, l, m and Extended Data Figs 1c, d, 5b), paired circular areas located between the eyes are interpreted as the nasal sacs. The proximity of these structures suggests a single median duct might have connected them. The paired eyes sometimes lie at a steep angle to the body axis (Extended Data Fig. 5b), suggesting that the head formed a discrete lobe capable of rotation. A median triangular area may represent cranial cartilage (Fig. 1d and Extended Data Fig. 4c–e), while darker dorsal elements might indicate possible arcualia (Fig. 1g).

Posterior to the head and ventrally positioned is a large branchial area. This region is identified on the basis of two sets of bars, seven in total, on each side of the body (Fig. 1d and Extended Data Fig. 4c). In dorso-ventral

¹Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK. ²Department of Natural History (Palaeobiology), Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada. ³University of Toronto, Department of Ecology and Evolutionary Biology, 25 Willcocks Street, Toronto, Ontario M5S 3B2, Canada.

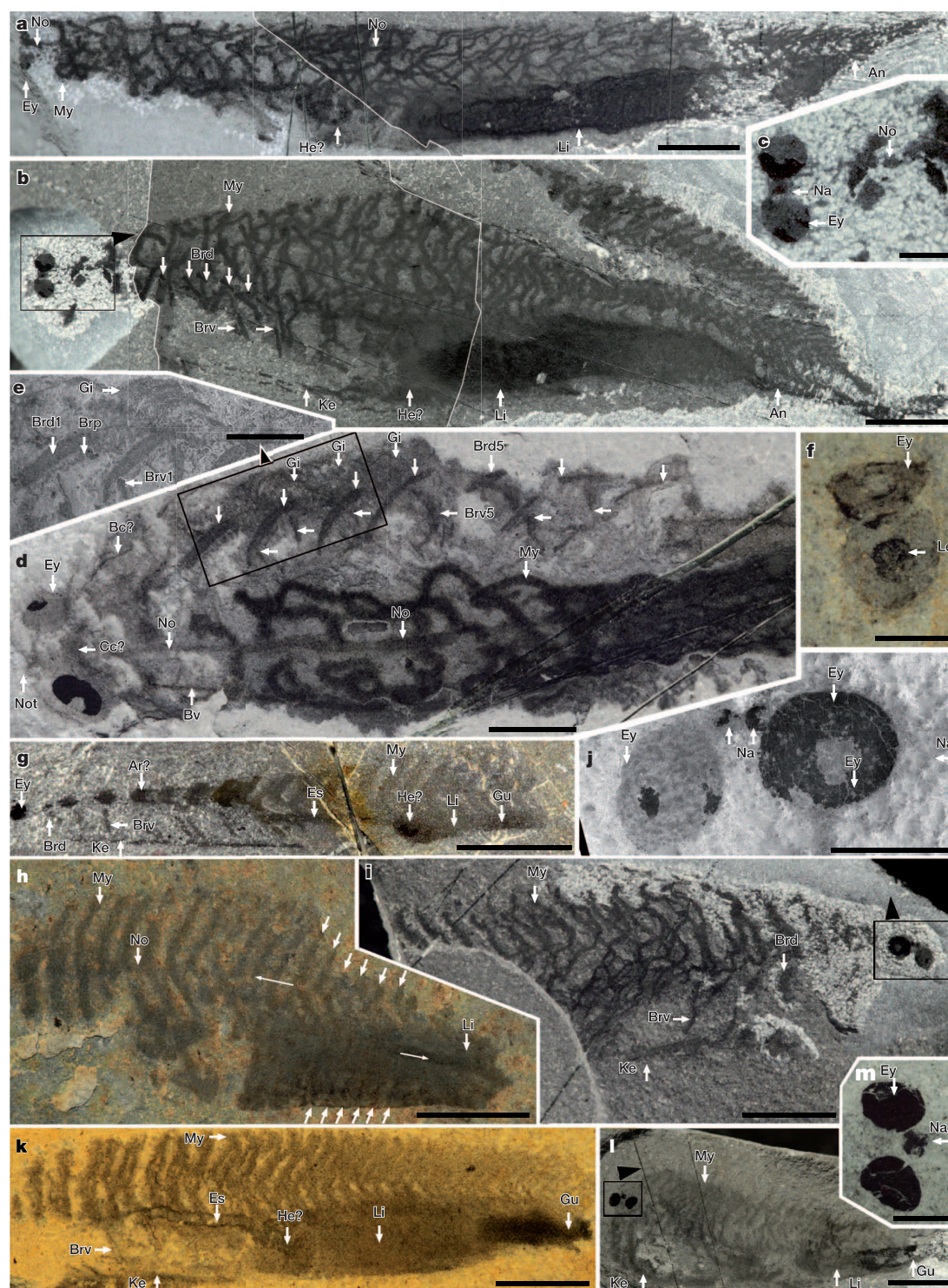


Figure 1 | *Metaspriggina walcotti* (Simonetta and Insom, 1993) specimens from Marble Canyon and Walcott's Quarry collected by the Royal Ontario Museum. **a–f, h–m**, Specimens collected from Marble Canyon. **g**, Specimens collected from Walcott's Quarry. **a–m**, Complete (**a–c**) and partial (**d–m**) specimens, anterior to the left (except **i**). **a**, ROM62935, oblique view. **b, c**, ROM62938, lateral view and close up of the anterior section (framed area in **c**). **d, e**, ROM62933, oblique view, and close up of gill bars (framed area in **e**). Only the position of some gill bars (Brd1 and Brv1 in **e**) are highlighted by arrows (see also Extended Data Fig. 4c). **f**, ROM62946, oblique view, pair of eyes showing presumed lens (see Extended Data Fig. 5d). **g**, ROM62964, lateral view. **h**, ROM62924, lateral view, showing W-shaped myomeres and main inflexions (thin arrows). **i, j**, ROM62932, lateral view, and close up showing

eyes (framed area in **j** with the image flipped vertically). **k**, ROM62954, lateral view showing internal organs. **l, m**, ROM62928, oblique view, and close up showing eyes in **m**. **a, b, e, h, j–m**, Backscatter scanning electron microscopy images (**e, j**), composite images of both parts and counterparts (**f–i, k–m**) and stitched images at white lines (**a, b**). An, anus; Ar?, possible arcualia; Bc?, possible extra-branchial cartilage; Brv, branchial bars (ventral element); Brd, branchial bars (dorsal element); Brp, branchial bar processes; Bv, blood vessel; Cc?, possible cranial cartilage; Es, oesophagus; Ey, eyes; Gi, gill filaments; Gu, gut; He?, possible heart; Ke, keel; Le, lens; Li, liver; My, myomere; Na, nasal sacs; No, notochord; Not, notch. Scale bars: 5 mm (**a, b, g–i, k, l**); 2 mm (**d**); 1 mm (**c, e, f, j, m**).

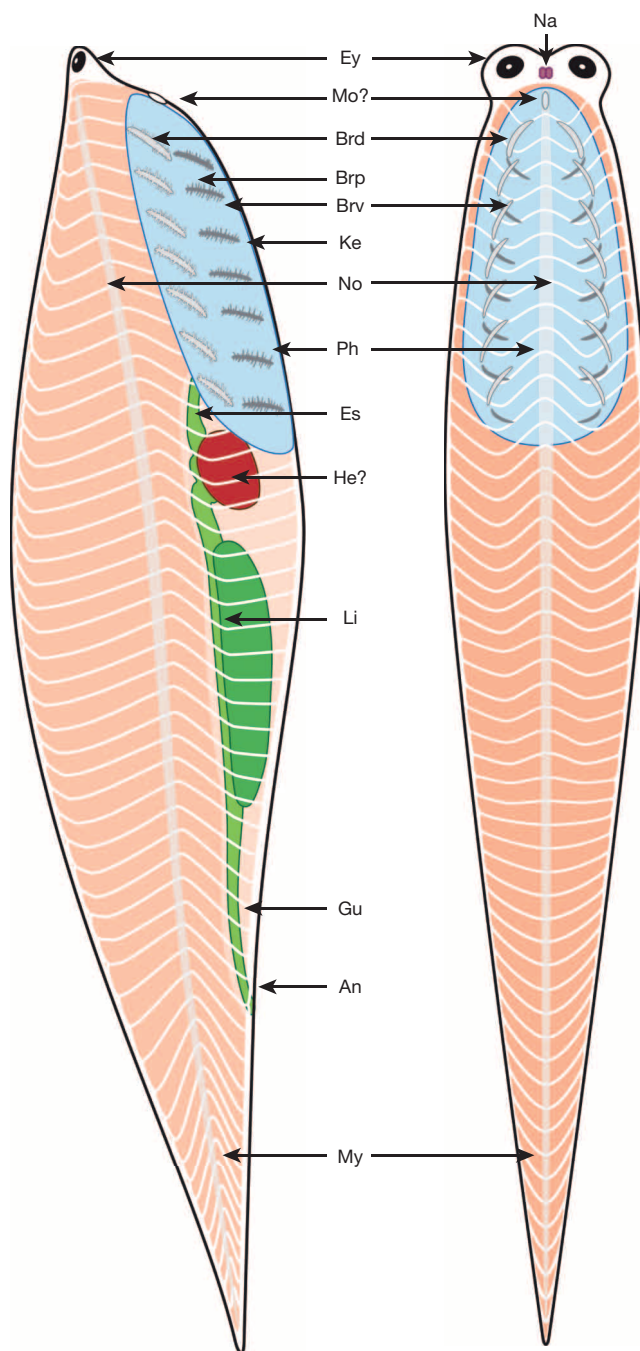


Figure 2 | *Metaspriggina walcotti* (Simonetta and Insom, 1993).

Diagrammatic anatomical drawing, dorsal and lateral views showing main morphological features (gills and possible blood vessels; arcualia and extra branchial and cranial cartilages have not been reconstructed). Mo?, possible position of mouth; Ph, pharyngeal area. For other abbreviations, see Fig. 1.

view both sets of bars are curved (Fig. 1d), while in lateral aspect the upper series is obliquely inclined ventrally and the lower series is more steeply angled (Fig. 1b, g). Bars may show a series of short extensions (Fig. 1d, e and Extended Data Fig. 4c). In addition, in one specimen (Fig. 1d, e) the anterior-most dorsal bar appears to be somewhat more robust (0.32 mm in thickness) than the other bars (average = 0.25 mm, standard deviation = 0.028). Closely spaced and relatively massive unequivocal gill filaments seem located distal to the associated bars (Fig. 1d, e), except perhaps along the anterior-most set of bars. No gill openings have been identified, but gills were possibly housed in pouches reminiscent of the arrangement in the related *Mylokunmingia*¹ and *Haikouichthys*³. Faint ventro-lateral structures parallel to the body (Fig. 1d) may represent extra-branchial cartilage.

From the roof of the posterior part of the branchial chamber a narrow strand, sometimes with positive relief, is interpreted as the oesophagus (Fig. 1g, k and Extended Data Fig. 5d, e) and this extends posteriorly towards the rest of the gut. An anterior area surrounding the gut trace possibly represents the heart (Fig. 1a, b, k and Extended Data Fig. 2e, 6b), followed by a darker area that may be a liver (Fig. 1a, b, g, h, k, l and Extended Data Figs 1a, g, 2b–e, 3b, c, 5a, c, e). Gut content is evident in the posterior section and just before the anus (Fig. 1a and Extended Data Fig. 1e, h). The latter is located on the posterior margin, thus defining a post-anal tail. Gut content is indeterminate, but *Metaspriggina* may have been microphagous with the short extensions on the gill bars possibly serving an analogous function to gill rakers.

Phylogenetic analysis (Fig. 3 and Supplementary Information) indicates *Metaspriggina* to be a stem vertebrate, possibly close to the slightly older Chengjiang taxa *Haikouichthys* and *Mylokunmingia* and otherwise basal to all other agnathans and gnathostomes. *Mylokunmingia* is known only from one specimen¹, but comparisons to *Haikouichthys*^{1–4} show similarities (for example, paired eyes, nasal sacs, lobate head), but also important differences. Although the myomeral configuration is also similar to *Haikouichthys*^{1,3}, in *Metaspriggina* the additional ventral chevron (Fig. 1h) and clear dorsal bend define a W-shaped arrangement directly comparable to fish¹⁶. The branchial region in *Metaspriggina* is also more voluminous, potentially consistent with greater respiratory demands. The branchial bars have a markedly different orientation from *Haikouichthys*, in which not only are they less well-preserved, but in contrast to *Metaspriggina* the presumed ceratobranchials are parallel to the ventral margin and the upper series of epibranchials lie at a steep angle^{1,2}.

The configuration of the branchial bars and associated gills in *Metaspriggina* has three wider implications. The first is that this arrangement recalls their frequently invoked ancestral configuration^{17–19}, even though

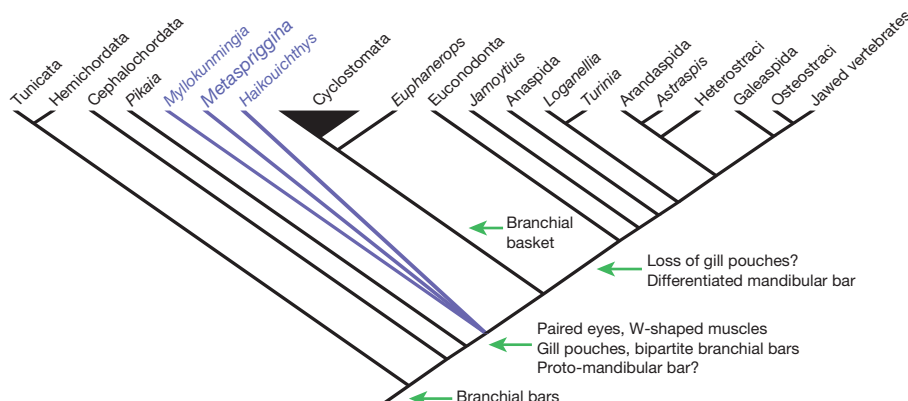


Figure 3 | Cladogram with backbone constraint for cyclostome monophyly, and using rescaled consistency indices, showing the position of *Metaspriggina* as part of basal stem-group soft-bodied vertebrates. The origin and potential loss of key vertebrate structures is indicated (see Supplementary Information for details).

this is usually regarded as hypothetical¹⁹. Primitive serial homology is often used as a starting point in discussions on the origin of jaws, although *Metaspriggina* is too basal (Fig. 3) to be directly informative in this regard²⁰. The second implication concerns possible links to the branchial configuration in extinct and extant lampreys. An intermediary stage between *Metaspriggina* and the complex branchial basket of living lamprey⁷ may occur in *Euphanerops*²¹ and the possibly synonymous *Endeiolepis*²². Although their unusual polybranchy may reflect hypoxia²², in these Late Devonian period (Frasnian) agnathans the branchial arches have fused into single sinuous units and are now associated with an array of small rods and 'copular elements'. This may prefigure²¹ the complex basket of extant lampreys⁷, a configuration that had apparently stabilized by the latest Devonian period (Famennian)²³.

Branchial arches of *Metaspriggina* also show two features reminiscent of gnathostomes. Although considered homologous^{8,24}, the position of the gills relative to the branchial arches differ, being interior in agnathans but exterior in the gnathostomes and apparently also in *Metaspriggina*. While differential preservation cannot be ruled out, the anterior-most branchial arch appears more robust and lacks gills. Correspondingly, in gnathostomes the mandibular arch is larger and clearly differentiated and only from the second branchial arch (hyoid) are gills borne⁸. Accordingly, if the anterior-most arch in *Metaspriggina* was to be equated with the mandibular arch then this might suggest that this region was always gill-less. In addition the putative extrabranchial cartilages of this Cambrian fish find a possible parallel in elasmobranchs^{8,24}, although convergence is perhaps more likely. Finally, both *Metaspriggina* and *Haikouichthys* are consistent with the earliest vertebrates possessing bipartite arches. Apart from a postulated fusion in the lamprey clade, it is suggested that unless homoplasious, this bipartite arrangement may have persisted crown-ward. If so, then models of jaw articulation such as 'hinges and caps'²⁵, invaluable in developmental studies, might need to be treated with caution²⁶ if employed in the context of atavistic homeotic transformations²⁷.

METHODS SUMMARY

Most specimens from Marble Canyon were collected *in situ* (Supplementary Table 1). Specimens were prepared mechanically to remove sediment concealing important morphological features. Lack of sediment fissility and difficulties of preparation often make exposure of the fossil (especially at either end) difficult. Specimens were studied using a range of photographic techniques commonly employed for this type of material⁶. Pictures of both parts and counterparts were merged in Adobe Photoshop CS6 using the 'apply image' function and 'darken' blending mode (Fig. 1a, b, f–i, k–m and Extended Data Figs 1–3, 5a, c, d). Scanning electron microscopy images were obtained using a JEOL JSM6610-Lv at the University of Toronto (Earth Sciences). Methods regarding phylogenetic analysis can be found in Supplementary Information.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 February; accepted 29 April 2014.

Published online 11 June 2014.

1. Shu, D.-G. *et al.* Lower Cambrian vertebrates from South China. *Nature* **402**, 42–46 (1999).
2. Shu, D.-G. *et al.* Head and backbone of the Early Cambrian vertebrate *Haikouichthys*. *Nature* **421**, 526–529 (2003).
3. Hou, X.-G., Aldridge, R. J., Siveter, D. J., Siveter, D. J. & Feng, X.-H. New evidence on the anatomy and phylogeny of the earliest vertebrates. *Proc. R. Soc. Lond. B* **269**, 1865–1869 (2002).
4. Zhang, X.-G. & Hou, X.-G. Evidence for a single median fin-fold and tail in the Lower Cambrian vertebrate, *Haikouichthys ercaicunensis*. *J. Evol. Biol.* **17**, 1162–1166 (2004).
5. Conway Morris, S. A redescription of a rare chordate, *Metaspriggina walcotti* Simonetta and Insom, from the Burgess Shale (Middle Cambrian), British Columbia, Canada. *J. Paleontol.* **82**, 424–430 (2008).

6. Caron, J.-B. *et al.* A new phyllopod bed-like assemblage from the Burgess Shale of the Canadian Rockies. *Nature Commun.* **5**, 3210 (2014).
7. Martin, W. M., Bumm, L. A. & McCauley, D. W. Development of the viscerocranial skeleton during embryogenesis of the sea lamprey, *Petromyzon marinus*. *Dev. Dyn.* **238**, 3126–3138 (2009).
8. Janvier, P. *Early Vertebrates* (Oxford Monographs on Geology and Geophysics) Vol. 33 (Clarendon, 1996).
9. Conway Morris, S. & Caron, J.-B. *Pikaia gracilens* Walcott, a stem-group chordate from the Middle Cambrian of British Columbia. *Biol. Rev. Camb. Philos. Soc.* **87**, 480–512 (2012).
10. Mallatt, J. & Holland, N. *Pikaia gracilens* Walcott: stem chordate, or already specialized in the Cambrian? *J. Exp. Zool. B Mol. Dev. Evol.* **320**, 247–271 (2013).
11. Sansom, R. S., Gabbott, S. E. & Purnell, M. A. Atlas of vertebrate decay: a visual and taphonomic guide to fossil interpretation. *Palaeontology* **56**, 457–474 (2013).
12. Johnston, K. J., Johnston, P. A. & Powell, W. G. A new Middle Cambrian, Burgess Shale-type biota, *Bolaspidella* Zone, Chancellor Basin, southeastern British Columbia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **277**, 106–126 (2009).
13. Webster, M. Systematic revision of the Cambrian trilobite *Bathynotus* Hall, 1860, with documentation of new occurrences in western Laurentia and implications for intercontinental biostratigraphic correlation. *Mem. Assoc. Australasian Palaeontologists* **37**, 369–406 (2009).
14. Conway Morris, S. Ediacaran-like fossils in Cambrian Burgess Shale-type faunas of North America. *Palaeontology* **36**, 593–635 (1993).
15. Lacalli, T. The Middle Cambrian fossil *Pikaia* and the evolution of chordate swimming. *EvoDevo* **3**, 12 (2012).
16. Van Leeuwen, J. L. A mechanical analysis of myomere shape in fish. *J. Exp. Biol.* **202**, 3405–3414 (1999).
17. Gillis, J. A., Modrell, M. S. & Baker, C. V. H. Developmental evidence for serial homology of the vertebrate jaw and gill arch skeleton. *Nature Commun.* **4**, 1436 (2013).
18. Kuratani, S. Evolution of the vertebrate jaw from developmental perspectives. *Evol. Dev.* **14**, 76–92 (2012).
19. Kuratani, S. Developmental studies of the lamprey and hierarchical evolutionary steps towards the acquisition of the jaw. *J. Anat.* **207**, 489–499 (2005).
20. Gai, Z.-K., Donoghue, P. C. J., Zhu, M., Janvier, P. & Stamparoni, M. Fossil jawless fish from China foreshadows early jawed vertebrate anatomy. *Nature* **476**, 324–327 (2011).
21. Janvier, P. & Arsenault, M. The anatomy of *Euphanerops longaeus* Woodward, 1900, an anaspid-like jawless vertebrate from the Upper Devonian of Miguasha, Quebec, Canada. *Geodiversitas* **29**, 143–216 (2007).
22. Janvier, P., Desbiens, S., Willett, J. A. & Arsenault, M. Lamprey-like gills in a gnathostome-related Devonian jawless vertebrate. *Nature* **440**, 1183–1185 (2006).
23. Gess, R. W., Coates, M. I. & Rubidge, B. S. A lamprey from the Devonian period of South Africa. *Nature* **443**, 981–984 (2006).
24. Mallatt, J. Early vertebrate evolution: pharyngeal structure and the origin of gnathostomes. *J. Zool.* **204**, 169–183 (1984).
25. Depew, M. J. & Compagnucci, C. Tweaking the hinge and caps: testing a model of the organization of jaws. *J. Exp. Zool. B Mol. Dev. Evol.* **310**, 315–335 (2008).
26. Medeiros, D. M. & Crump, J. G. New perspectives on pharyngeal dorsoventral patterning in development and evolution of the vertebrate jaw. *Dev. Biol.* **371**, 121–135 (2012).
27. Nichols, J. T., Pan, L.-Y., Moens, C. B. & Kimmel, C. B. *barx1* represses joints and promotes cartilage in the craniofacial skeleton. *Development* **140**, 2765–2775 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Lacalli for comments on an earlier draft of the manuscript, J. Mallatt for an extensive series of critiques, and J. Hoyal-Cuthill and C. Aria for assistance with phylogenetic analyses. We also thank M. Collins for technical drawings and reconstructions, and R. Thomas and M. Webster for information on the Kinzers and Parker formations, respectively. S.C.M. thanks V. Brown for manuscript preparation, and the Department of Earth Sciences and St John's College, Cambridge for support. We thank Parks Canada for granting a collection and research permit to J.-B.C. (YNP-2012-12054). Fieldwork support for the 2012 expedition comes from the Royal Ontario Museum (DMV Research and Acquisition Fund and DNH Fieldwork Fund), M. Streng (Uppsala University and the Swedish Research Council), R. Gaines (Pomona College), G. Mangano (University of Saskatchewan) and a Natural Sciences and Engineering Research Council Discovery Grant (to J.-B.C., #341944). This is Royal Ontario Museum Burgess Shale project number 53.

Author Contributions J.-B.C. collected fossils, prepared all illustrative material and conducted phylogenetic analyses. S.C.M. wrote early drafts of paper, and both authors discussed results and developed observations and conclusions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C.M. (sc113@cam.ac.uk) or J.-B.C. (jcaron@rom.on.ca).

Neural constraints on learning

Patrick T. Sadtler^{1,2,3}, Kristin M. Quick^{1,2,3}, Matthew D. Golub^{2,4}, Steven M. Chase^{2,5}, Stephen I. Ryu^{6,7}, Elizabeth C. Tyler-Kabara^{1,8,9}, Byron M. Yu^{2,4,5*} & Aaron P. Batista^{1,2,3*}

Learning, whether motor, sensory or cognitive, requires networks of neurons to generate new activity patterns. As some behaviours are easier to learn than others^{1,2}, we asked if some neural activity patterns are easier to generate than others. Here we investigate whether an existing network constrains the patterns that a subset of its neurons is capable of exhibiting, and if so, what principles define this constraint. We employed a closed-loop intracortical brain–computer interface learning paradigm in which Rhesus macaques (*Macaca mulatta*) controlled a computer cursor by modulating neural activity patterns in the primary motor cortex. Using the brain–computer interface paradigm, we could specify and alter how neural activity mapped to cursor velocity. At the start of each session, we observed the characteristic activity patterns of the recorded neural population. The activity of a neural population can be represented in a high-dimensional space (termed the neural space), wherein each dimension corresponds to the activity of one neuron. These characteristic activity patterns comprise a low-dimensional subspace (termed the intrinsic manifold) within the neural space. The intrinsic manifold presumably reflects constraints imposed by the underlying neural circuitry. Here we show that the animals could readily learn to proficiently control the cursor using neural activity patterns that were within the intrinsic manifold. However, animals were less able to learn to proficiently control the cursor using activity patterns that were outside of the intrinsic manifold. These results suggest that the existing structure of a network can shape learning. On a timescale of hours, it seems to be difficult to learn to generate neural activity patterns that are not consistent with the existing network structure. These findings offer a network-level explanation for the observation that we are more readily able to learn new skills when they are related to the skills that we already possess^{3,4}.

Some behaviours are easier to learn than others^{1–4}. We hypothesized that the ease or difficulty with which an animal can learn a new behaviour is determined by the current properties of the networks of neurons governing the behaviour. We tested this hypothesis in the context of brain–computer interface (BCI) learning. In a BCI paradigm, the user controls a cursor on a computer screen by generating activity patterns across a population of neurons. A BCI offers advantages for studying learning because we can observe all of the neurons that directly control an action, and we can fully specify the mapping from neural activity to action. This allowed us to define which activity patterns would lead to task success and to test whether subjects were capable of generating them. Previous studies have shown that BCI learning can be remarkably extensive^{5–10}, raising the intriguing possibility that most (or all) novel BCI mappings are learnable.

Two male Rhesus macaques (aged 7 and 8 years) were trained to move a cursor from the centre of a screen to one of eight radially arranged targets by modulating the activity of 85–91 neural units (that is, threshold crossings on each electrode) recorded in the primary motor cortex (Fig. 1a). To represent the activity of the neural population, we defined a high-dimensional space (called the neural space) where each axis corresponds

to the activity of one neural unit. The activity of all neural units during a short time period is represented as a point in this space (Fig. 1b). At each time step, the neural activity (a green point in Fig. 1b) is mapped onto a control space (black line in Fig. 1b; two-dimensional plane in the actual experiments, corresponding to horizontal and vertical cursor velocity) to specify cursor velocity. The control space is the geometrical representation of a BCI mapping. At the start of each day, we calibrated an ‘intuitive mapping’ by specifying a control space that the monkey used to move the cursor proficiently (Extended Data Fig. 1).

At the beginning of each day we also characterized how each neural unit changed its activity relative to the other neural units (that is, how the neural units co-modulated). In the simplified network represented in Fig. 1b, neurons 1 and 3 positively co-modulate due to common input, whereas neurons 1 and 2 negatively co-modulate due to an indirect inhibitory connection. Such co-modulations among neurons mean that neural activity does not uniformly populate the neural space^{11–16}. We identified the low-dimensional space that captured the natural patterns of co-modulation among the recorded neurons. We refer to this space as the intrinsic manifold (yellow plane in Fig. 1b, c). By construction, the intuitive mapping lies within the intrinsic manifold. Our key experimental manipulation was to change the BCI mapping so that the control space was either within or outside of the intrinsic manifold. A within-manifold perturbation was created by re-orienting the intuitive control space but keeping it within the intrinsic manifold (depicted as the red line in Fig. 1c). This preserved the relationship between neural units and co-modulation patterns, but it altered the way in which co-modulation patterns affected cursor kinematics (red arrows, Fig. 1a). An outside-manifold perturbation was created by re-orienting the intuitive control space and allowing it to depart from the intrinsic manifold (depicted as the blue line in Fig. 1c). This altered the way in which neural units contributed to co-modulation patterns, but it preserved the way in which co-modulation patterns affected cursor kinematics (blue arrows, Fig. 1a). In both cases, performance was impaired once the new mapping was introduced, and we observed whether the monkeys could learn to regain proficient control of the cursor.

To regain proficient control of the cursor under a within-manifold perturbation, the animals had to learn new associations between the natural co-modulation patterns and the cursor kinematics (Fig. 1d). To restore proficient control of the cursor under an outside-manifold perturbation, the animals had to learn to generate new co-modulation patterns among the recorded neurons. Our hypothesis predicted that within-manifold perturbations would be more readily learnable than outside-manifold perturbations.

Just after the perturbed mappings were introduced, BCI performance was impaired (Fig. 2a, b, first grey vertical band). Performance improved for the within-manifold perturbation (Fig. 2a), showing that the animal learned to control the cursor under that mapping. In contrast, performance remained impaired for the outside-manifold perturbation (Fig. 2b), showing that learning did not occur. We quantified the amount

¹Department of Bioengineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. ²Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania 15213, USA. ³Systems Neuroscience Institute, University of Pittsburgh, Pittsburgh Pennsylvania 15261, USA. ⁴Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.

⁵Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. ⁶Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA.

⁷Department of Neurosurgery, Palo Alto Medical Foundation, Palo Alto, California 94301, USA. ⁸Department of Physical Medicine and Rehabilitation, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁹Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA.

*These authors contributed equally to this work.

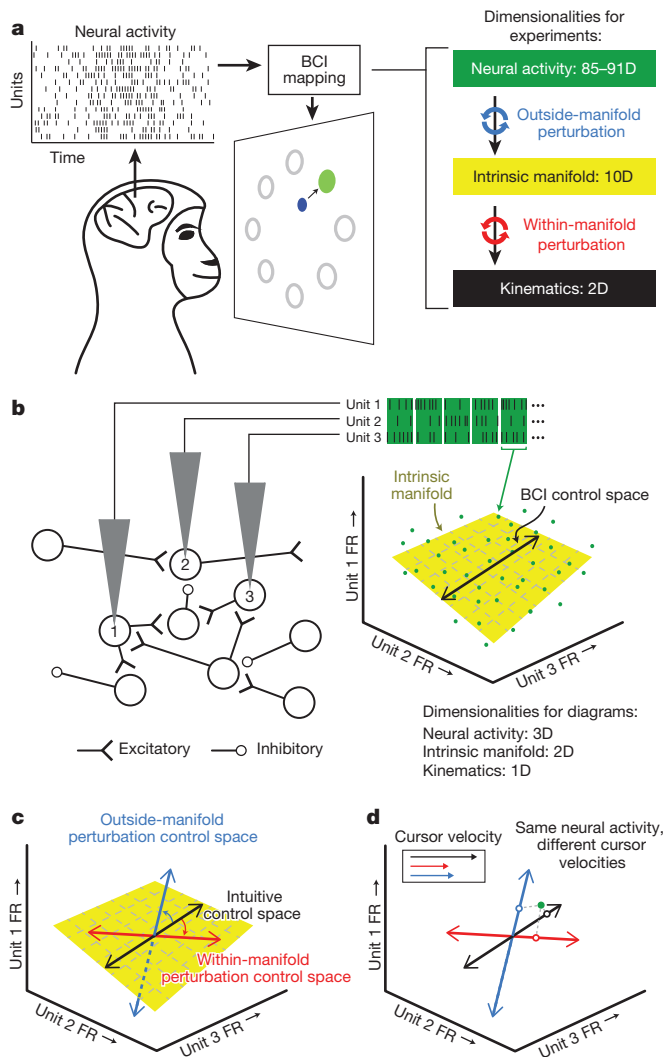


Figure 1 | Using a brain-computer interface to study learning. **a**, Monkeys moved the BCI cursor (blue circle) to acquire targets (green circle) by modulating their neural activity. The BCI mapping consisted of first mapping the population neural activity to the intrinsic manifold using factor analysis, then from the intrinsic manifold to cursor kinematics using a Kalman filter. This two-step procedure allowed us to perform outside-manifold perturbations (blue arrows) and within-manifold perturbations (red arrows). **b**, A simplified, conceptual illustration using three electrodes. The firing rate (FR) observed on each electrode in a brief epoch define a point (green dots) in the neural space. The intrinsic manifold (yellow plane) characterizes the prominent patterns of co-modulation. Neural activity maps onto the control space (black line) to specify cursor velocity. **c**, Control spaces for an intuitive mapping (black arrow), within-manifold perturbation (red arrow) and outside-manifold perturbation (blue arrow). **d**, Neural activity (green dot) elicits different cursor velocities (open circles and inset) under different mappings. Arrow colours as in **c**.

of learning as the extent to which BCI performance recovered from its initial impairment to the level attained while using the intuitive mapping (Fig. 2c). For within-manifold perturbations, the animals regained proficient control of the cursor (red histograms in Fig. 2d and Extended Data Fig. 2), indicating that they could learn new associations between natural co-modulation patterns and cursor kinematics. For outside-manifold perturbations, BCI performance remained impaired (blue histograms in Fig. 2d and Extended Data Fig. 2), indicating that it was difficult to learn to generate new co-modulation patterns, even when those patterns would have led to improved performance in the task. These results support our hypothesis that the structure of a network

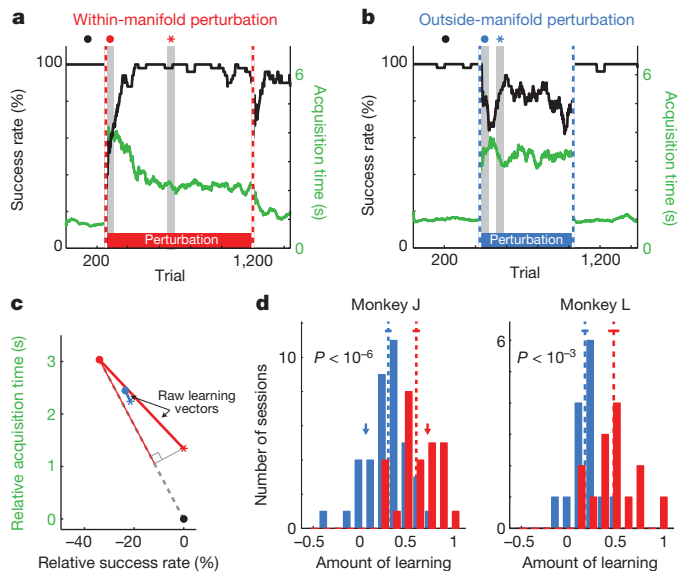


Figure 2 | Better learning for within-manifold perturbations than outside-manifold perturbations. **a**, **b**, Task performance during one representative within-manifold perturbation session (**a**) and one representative outside-manifold perturbation session (**b**). Black trace, success rate; green trace, target acquisition time. Dashed vertical lines indicate when the BCI mapping changed. Grey vertical bands represent 50-trial bins used to determine initial (red and blue dots) and best (red and blue asterisks) performance with the perturbed mapping. **c**, Quantifying the amount of learning. Black dot, performance with the intuitive mappings; red and blue dots, performance (success rate and acquisition time are relative to performance with intuitive mapping) just after the perturbation was introduced for sessions in Fig. 2a and Fig. 2b; red and blue asterisks, best performance during those perturbation sessions; dashed line, maximum learning vector for the session in Fig. 2a. The amount of learning for each session is the length of the raw learning vector projected onto the maximum learning vector, normalized by the length of the maximum learning vector. This is the ratio of the length of the thin red line to the length of the dashed line. **d**, Amount of learning for all sessions. A value of 1 indicates complete learning of the relationship between neural activity and kinematics, and 0 indicates no learning. Learning is significantly better for within-manifold perturbations (red, $n = 28$ (monkey J), 14 (monkey L)) than for outside-manifold perturbations (blue, $n = 39$ (monkey J), 15 (monkey L)). Arrows indicate the sessions shown in Fig. 2a (red) and Fig. 2b (blue). Dashed lines, means of distributions; solid lines, mean \pm standard error of the mean (s.e.m.). P values were obtained from two-tailed Student's t -tests.

determines which patterns of neural activity (and corresponding behaviours) a subject can readily learn to generate.

Two additional lines of evidence show that BCI control was more learnable when using within-manifold perturbations than outside-manifold perturbations. First, perturbation types differed in their after-effects. After a lengthy exposure to the perturbed mapping, we again presented the intuitive mapping (the second dashed vertical line in Fig. 2a, b). Following within-manifold perturbations, performance was impaired briefly (Extended Data Fig. 3, red histogram), indicating that learning had occurred¹⁷. Following outside-manifold perturbations, performance was not impaired, which is consistent with little, if any, learning having occurred (Extended Data Fig. 3, blue histogram). Second, the difference in learnability between the two types of perturbation was present from the earliest sessions, and over the course of the study the monkeys did not improve at learning (Extended Data Fig. 4).

These results show that the intrinsic manifold was a reliable predictor of the learnability of a BCI mapping: new BCI mappings that were within the intrinsic manifold were more learnable than those outside of it. We considered five alternative explanations for the difference in learnability. First, we considered the possibility that mappings which were more difficult to use initially might be more difficult to learn. We ensured that the initial performance impairments were equivalent for the two perturbation types (Fig. 3a).

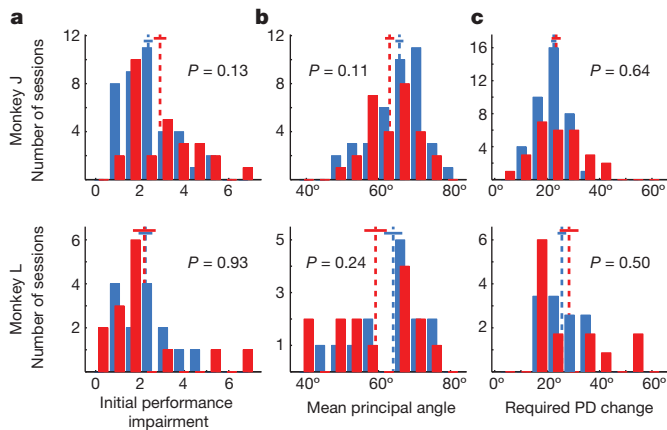


Figure 3 | Alternative explanations do not explain the difference in learnability between the two types of perturbation. **a**, Performance impairment immediately following within-manifold and outside-manifold perturbations. **b**, Mean principal angles between intuitive and perturbed mappings. **c**, Mean required change in preferred direction (PD) for individual neural units. For all panels: red, within-manifold perturbations; blue, outside-manifold perturbations; dashed lines, means of distributions; solid lines, mean \pm s.e.m.; P values are for two-tailed Student's t -tests; same number of sessions as in Fig. 2d.

Second, we posited that the animals must search through neural space for the new control space following the perturbation. If the control spaces for one type of perturbation tended to be farther from the intuitive control space, then they might be harder to find, and thus, learning would be reduced. We ensured that the angles between the intuitive and perturbed control spaces did not differ between the two perturbation types (Fig. 3b). Incidentally, Fig. 3b also shows that the perturbations were not pure workspace rotations. If that were the case, the angles between control spaces would have been zero, not in the range of 40° – 80° as shown.

Third, we considered how much of an impact the perturbations exerted on the activity of each neural unit. Learning is manifested (at least in part) as changes in the preferred direction (that is, the direction of movement for which a neuron is most active) of individual neurons^{7,18}. If learning one type of perturbation required larger changes in preferred directions of neural units, then those perturbations might be harder to learn. We predicted the changes in preferred directions that would be required to learn each perturbation while minimizing changes in activity. We ensured that learning the two perturbation types required comparable preferred-direction changes (Fig. 3c).

Fourth, for one monkey (L), we ensured that the sizes of the search spaces for finding a strategy to proficiently control the cursor were the same for both perturbation types (see Methods).

Fifth, hand movements were comparable and nearly non-existent for both perturbation types and should therefore have had no impact on learnability (Extended Data Fig. 5).

We conclude from these analyses that the parsimonious explanation for BCI learning is whether or not the new control space is within the intrinsic manifold. These alternative explanations did reveal interesting secondary aspects of the data; they partially explained within-category differences in learnability, albeit in an idiosyncratic manner between the two monkeys (Extended Data Fig. 6).

A key step in these experiments was the identification of an intrinsic manifold using dimensionality reduction¹¹. Although our estimate of the intrinsic manifold can depend on several methodological factors (Extended Data Fig. 7 caption), the critical property of such a manifold is that it captures the prominent patterns of co-modulation among the recorded neurons, which presumably reflect underlying network constraints. For consistency, we estimated a linear, ten-dimensional intrinsic manifold each day. In retrospect, we considered whether our choice of ten dimensions had been appropriate (Fig. 4). We estimated the intrinsic dimensionality of the neural activity for each day (Fig. 4a); the average

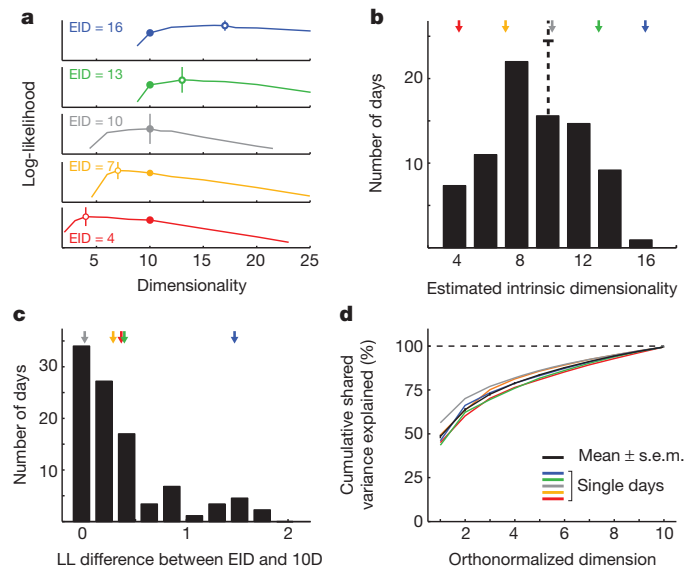


Figure 4 | Properties of the intrinsic manifold. **a**, Cross-validated log-likelihoods (LL; arbitrary units) of the population activity for different days. The peaks (open circles) indicate the estimated intrinsic dimensionality (EID). Vertical bars indicate the standard error of log-likelihoods, computed across four cross-validation folds. A ten-dimensional intrinsic manifold was used for all experiments (solid circles). **b**, EID across all days and both monkeys (mean \pm s.e.m., 9.81 ± 0.31). **c**, Difference between the LL for the ten-dimensional (10D) model and the EID model. Units are the number of standard errors of LL for the EID model. For 89% (78 of 88) of the days, the LL for the ten-dimensional model was within one standard error of the EID model. All sessions were less than two standard errors away. **d**, Cumulative shared variance explained by the ten-dimensional intrinsic manifold used during the experiment. Coloured curves correspond to the experimental days shown in Fig. 4a. The black curve shows the mean \pm s.e.m. across all days ($n = 88$; monkey J, 58; monkey L, 30).

dimensionality was about ten (Fig. 4b). Even though the estimated dimensionalities ranged from 4–16, the selection of ten dimensions still provided a model that was nearly as good as the best model (Fig. 4c). Because the top few dimensions captured the majority of the co-modulation among the neural units (Fig. 4d), we probably could have selected a different dimensionality within the range of near-optimal dimensionalities and still attained similar results (Extended Data Fig. 7 caption). We note that we cannot make claims about the ‘true’ dimensionality of the primary motor cortex, in part because it probably depends on considerations such as the behaviours the animal is performing and, perhaps, its level of skill.

Sensorimotor learning probably encompasses a variety of neural mechanisms, operating at diverse timescales and levels of organization. We posit that learning a within-manifold perturbation harnesses the fast-timescale learning mechanisms that underlie adaptation¹⁹, whereas learning an outside-manifold perturbation engages the neural mechanisms required for skill learning^{20,21}. This suggests that learning outside-manifold perturbations could benefit from multi-day use^{5,22}. Such learning might require the intrinsic manifold to expand or change orientation.

Other studies have employed dimensionality-reduction techniques to interpret how networks of neurons encode information^{11–16} and change their activity during learning^{23,24}. Our findings strengthen those discoveries by showing that low-dimensional projections of neural data are not only visualization tools—they can reveal causal constraints on the activity attainable by networks of neurons. Our study also indicates that the low-dimensional patterns present among a population of neurons may better reflect the elemental units of volitional control than do individual neurons.

In summary, a BCI paradigm enabled us to reveal neural constraints on learning. The principles we observed may govern other forms of learning^{4,25–28} and perhaps even cognitive processes. For example, combinatorial creativity²⁹, which involves re-combining cognitive elements in

new ways, might involve the generation of new neural activity patterns that are within the intrinsic manifold of relevant brain areas. Transformational creativity, which involves creating new cognitive elements, may result from generating neural activity patterns outside of the relevant intrinsic manifold. More broadly, our results help to provide a neural explanation for the balance we possess between adaptability and persistence in our actions and thoughts³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 February; accepted 7 July 2014.

- Krakauer, J. W. & Mazzoni, P. Human sensorimotor learning: adaptation, skill, and beyond. *Curr. Opin. Neurobiol.* **21**, 636–644 (2011).
- Ranganathan, R., Wieser, J., Mosier, K. M., Mussa-Ivaldi, F. A. & Scheidt, R. A. Learning redundant motor tasks with and without overlapping dimensions: facilitation and interference effects. *J. Neurosci.* **34**, 8289–8299 (2014).
- Thoroughman, K. & Taylor, J. Rapid reshaping of human motor generalization. *J. Neurosci.* **25**, 8948–8953 (2005).
- Braun, D., Mehring, C. & Wolpert, D. Structure learning in action. *Behav. Brain Res.* **206**, 157–165 (2010).
- Ganguly, K. & Carmena, J. M. Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol.* **7**, e1000153 (2009).
- Fetz, E. E. Operant conditioning of cortical unit activity. *Science* **163**, 955–958 (1969).
- Jarosiewicz, B. et al. Functional network reorganization during learning in a brain-computer interface paradigm. *Proc. Natl Acad. Sci. USA* **105**, 19486–19491 (2008).
- Hwang, E. J., Bailey, P. M. & Andersen, R. A. Volitional control of neural activity relies on the natural motor repertoire. *Curr. Biol.* **23**, 353–361 (2013).
- Rouse, A. G., Williams, J. J., Wheeler, J. J. & Moran, D. W. Cortical adaptation to a chronic micro-electrocorticographic brain computer interface. *J. Neurosci.* **33**, 1326–1330 (2013).
- Engelhard, B., Ozeri, N., Israel, Z., Bergman, H. & Vaadia, E. Inducing γ oscillations and precise spike synchrony by operant conditioning via brain-machine interface. *Neuron* **77**, 361–375 (2013).
- Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.3776>.
- Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).
- Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
- Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
- Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
- Luczak, A., Barthó, P. & Harris, K. D. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**, 413–425 (2009).
- Shadmehr, R., Smith, M. & Krakauer, J. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108 (2010).
- Li, C. S., Padoa-Schioppa, C. & Bizzi, E. Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron* **30**, 593–607 (2001).
- Salinas, E. Fast remapping of sensory stimuli onto motor actions on the basis of contextual modulation. *J. Neurosci.* **24**, 1113–1118 (2004).
- Picard, N., Matsuzaka, Y. & Strick, P. L. Extended practice of a motor skill is associated with reduced metabolic activity in M1. *Nature Neurosci.* **16**, 1340–1347 (2013).
- Riout-Pedotti, M.-S., Friedman, D. & Donoghue, J. P. Learning-induced LTP in neocortex. *Science* **290**, 533–536 (2000).
- Peters, A. J., Chen, S. X. & Komiyama, T. Emergence of reproducible spatiotemporal activity during motor learning. *Nature* **510**, 263–267 (2014).
- Paz, R., Natan, C., Boraud, T., Bergman, H. & Vaadia, E. Emerging patterns of neuronal responses in supplementary and primary motor areas during sensorimotor adaptation. *J. Neurosci.* **25**, 10941–10951 (2005).
- Durstewitz, D., Vitoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66**, 438–448 (2010).
- Jeanne, J. M., Sharpee, T. O. & Gentner, T. Q. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**, 352–363 (2013).
- Gu, Y. et al. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011).
- Ingvalson, E. M., Holt, L. L. & McClelland, J. L. Can native Japanese listeners learn to differentiate /r-/ on the basis of F3 onset frequency? *Biling. Lang. Cogn.* **15**, 255–274 (2012).
- Park, D. C. et al. The impact of sustained engagement on cognitive function in older adults: the Synapse Project. *Psychol. Sci.* **25**, 103–112 (2014).
- Boden, M. A. Creativity and artificial intelligence. *Artif. Intell.* **103**, 347–356 (1998).
- Ajemian, R., D'Ausilio, A., Moorman, H. & Bizzi, E. A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits. *Proc. Natl Acad. Sci. USA* **110**, E5078–E5087 (2013).

Acknowledgements We thank A. Barth, C. Olson, D. Sussillo, R. J. Tibshirani and N. Urban for discussions; S. Flesher for help with data collection and R. Dum for advice on array placement. This work was funded by NIH NICHD CRCNS R01-HD071686 (A.P.B. and B.M.Y.), NIH NINDS R01-NS065065 (A.P.B.), Burroughs Wellcome Fund (A.P.B.), NSF DGE-0549352 (P.T.S.) and NIH P30-NS076405 (Systems Neuroscience Institute).

Author Contributions P.T.S., K.M.Q., M.D.G., S.M.C., B.M.Y. and A.P.B. designed the experiments. S.I.R. and E.C.T.-K. implanted the arrays. P.T.S. collected and analysed the data. P.T.S., B.M.Y. and A.P.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.M.Y. (byronyu@cmu.edu) or A.P.B. (apb10@pitt.edu).

METHODS

Electrophysiology and behavioural monitoring. We recorded from the proximal arm region of the primary motor cortex in two male Rhesus macaques (*Macaca mulatta*, aged 7 and 8 years) using 96-channel microelectrode arrays (Blackrock Microsystems) as the monkeys sat head-fixed in a primate chair. All animal handling procedures were approved by the University of Pittsburgh Institutional Animal Care and Use Committee. At the beginning of each session, we estimated the root-mean-square voltage of the signal on each electrode while the monkeys sat calmly in a darkened room. We then set the spike threshold at 3.0 times the root-mean-square value for each channel. Spike counts used for BCI control were determined from the times at which the voltage crossed this threshold. We refer to the threshold crossings recorded on one electrode as one neural unit. We used 85–91 neural units each day. We did not use an electrode if the threshold crossing waveforms did not resemble action potentials or if the electrode was electrically shorted to another electrode. The data were recorded approximately 19–24 months after array implantation for monkey J and approximately 8–9 months after array implantation for monkey L.

We monitored hand movements using an LED marker (PhaseSpace Inc.) on the hand contralateral to the recording array. The monkeys' arms were loosely restrained. The monkeys could have moved their forearms by approximately 5 cm from their arm rests, and there were no restrictions on wrist movement. The hand movements during the BCI trials were minimal, and we observed that the monkeys' movements did not approach the limits of the restraints. Extended Data Fig. 5a shows the average hand speed during the BCI trials. For comparison, Extended Data Fig. 5b shows the average hand speed during a standard point-to-point reaching task. We also recorded the monkeys' gaze direction (SR Research Ltd). Those data are not analysed here.

Task flow. Each day began with a calibration block during which we determined the parameters of the intuitive mapping. The monkeys then used the intuitive mapping for 400 trials (monkey J) or 250 trials (monkey L) during the baseline block. We then switched to the perturbed mapping for 600 trials (monkey J) or 400 trials (monkey L) for the perturbation block. This was followed by a 200-trial washout block with the intuitive mapping. Together, the perturbation and washout blocks comprised a perturbation session. The transitions between blocks were made seamlessly, without an additional delay between trials. We gave the monkey no indication which type of perturbation would be presented. On most days, we completed one perturbation session (monkey J, 50 of 58 days; monkey L, 29 of 30 days). On nine days, we completed multiple perturbation sessions.

Experimental sessions. We conducted 78 (30 within-manifold perturbations; 48 outside-manifold perturbations) sessions with monkey J. We conducted 31 sessions (16 within-manifold perturbations; 15 outside-manifold perturbations) with monkey L. For both monkeys, we did not analyse a session if the monkey attempted fewer than 100 trials with the perturbed mapping. For monkey J, we did not analyse 11 sessions (2 within-manifold perturbations; 9 outside-manifold perturbations). For monkey L, we did not analyse 3 sessions (2 within-manifold perturbations; 1 outside-manifold perturbation).

BCI calibration procedures. Each day began with a calibration block of trials. The data that we recorded during these blocks were used to estimate the intrinsic manifold and to calibrate the parameters of the intuitive mappings. For monkey J, we used two calibration methods (only one on a given day), and for monkey L, we used one method for all days.

The following describes the BCI calibration procedures for monkey J. The first method for this monkey relied on the neural signals being fairly stable across days. At the beginning of each day, the monkey was typically able to control the cursor proficiently using the previous day's intuitive mapping. We collected data for calibration by having the monkey use the previous day's intuitive mapping for 80 trials (10 per target).

We designed the second method because we were concerned about the potential for carry-over effects across days. This method relied on passive observation of cursor movement³¹. The monkey observed the cursor automatically complete the centre-out task for 80 trials (10 per target). At the beginning of each trial, the cursor appeared in the centre of the monkey's workspace for 300 ms. Then, the cursor moved at a constant velocity (0.15 m s^{-1}) to the pseudo-randomly chosen target for each trial. When the cursor reached the target, the monkey received a juice reward. After each trial, there was a blank screen for 200 ms before the next trial.

For both methods for monkey J, we used the neural activity recorded 300 ms after the start of each trial until the cursor reached the peripheral target for BCI calibration.

The following describes the BCI calibration procedure for monkey L. We observed that neural activity for this monkey was not as stable from day to day as it was for monkey J. As a result, we could not use the calibration procedure relying on the previous day's intuitive mapping. Additionally, the observation-based calibration procedure was not as effective at generating an intuitive decoder for monkey L as it had been for monkey J. Therefore, we used a closed-loop calibration procedure

(similar to reference 32) to generate the intuitive decoder. The procedure began with 16 trials (2 to each target) of the observation task. We calibrated a decoder from these 16 trials in the same manner as the first method for monkey J. We then switched to the BCI centre-out task, and the monkey controlled the velocity of the cursor using the decoder calibrated on the 16 observation trials. We restricted movement of the cursor so that it moved in a straight line towards the target (that is, any cursor movement perpendicular to the straight path to the target was scaled by a factor of 0). After 8 trials (1 to each target), we calibrated another decoder from those 8 trials. The monkey then controlled the cursor for 8 more trials with this newly calibrated decoder with perpendicular movements scaled by a factor of 0.125. We then calibrated a new decoder using all 16 closed-loop trials. We repeated this procedure over a total of 80 trials until the monkey was in full control of the cursor (perpendicular velocity scale factor = 1). We calibrated the intuitive mapping using the 80 trials during which the monkey had full or partial control of the cursor. For each of those trials, we used the neural activity recorded 300 ms after the start of the trial until the cursor reached the peripheral target.

BCI centre-out task. The same closed-loop BCI control task was used during the baseline, perturbation and washout blocks. At the beginning of each trial, the cursor (circle, radius = 18 mm) appeared in the centre of the workspace. One of eight possible peripheral targets (chosen pseudo-randomly) was presented (circle, radius = 20 mm; 150 mm (monkey J) or 125 mm (monkey L) from centre of workspace, separated by 45°). A 300 ms freeze period ensued, during which the cursor did not move. After the freeze period, the velocity of the cursor was controlled by the monkey through the BCI mapping. The monkey had 7,500 ms to move the cursor into the peripheral target. If the cursor acquired the peripheral target within the time limit, the monkey received a juice reward. After 200 ms, the next trial began. With the intuitive mappings, the monkeys' movement times were near 1,000 ms (Extended Data Fig. 1), but the monkeys sometimes exceeded the 7,500 ms acquisition time limit with the perturbed mappings. If the cursor did not acquire the target within the time limit, there was a 1,500 ms time-out before the start of the next trial.

Estimation of the intrinsic manifold. We identified the intrinsic manifold from the population activity recorded during the calibration session using the dimensionality reduction technique factor analysis^{33,34}. The central idea is to describe the high-dimensional population activity \mathbf{u} in terms of a low-dimensional set of factors \mathbf{z} . Each factor is distributed according to the standard normal distribution N . This can be written in vector form as:

$$\mathbf{z} \sim N(\mathbf{0}, I) \quad (1)$$

where I is the identity matrix. The neural activity is related to those factors by:

$$\mathbf{u}|\mathbf{z} \sim N(\Lambda\mathbf{z} + \boldsymbol{\mu}, \boldsymbol{\psi}) \quad (2)$$

where $\mathbf{u} \in \mathbb{R}^{q \times 1}$ is a vector of z -scored spike counts (z -scoring was performed separately for each neural unit) taken in non-overlapping 45 ms bins across the q neural units, and $\mathbf{z} \in \mathbb{R}^{10 \times 1}$ contains the ten factors. That is, the neural activity \mathbf{u} given a set of factors \mathbf{z} is distributed according to a normal distribution with mean $\Lambda\mathbf{z} + \boldsymbol{\mu}$ and diagonal covariance $\boldsymbol{\psi}$. The intrinsic manifold is defined as the column space of Λ . Each factor, or latent dimension, is represented by a column of Λ . We estimated Λ , $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$ using the expectation-maximization algorithm³⁵. The data collected during the calibration sessions had $1,470 \pm 325$ (monkey J, mean \pm standard deviation) and $1,379 \pm 157$ (monkey L) samples.

Intuitive mappings. The intuitive mapping was a modified version of the standard Kalman filter³⁶. A key component of the experimental design was to use the Kalman filter to relate factors (\mathbf{z}) to cursor kinematics rather than to relate neural activity directly to the cursor kinematics. This modification allowed us to perform the two different types of perturbation. We observed that performance with our modified Kalman filter is qualitatively similar to performance with a standard Kalman filter (data not shown).

The first step in the construction of the intuitive mapping was to estimate the factors using factor analysis (equations (1) and (2)). For each z -scored spike count vector \mathbf{u}_t , we computed the posterior mean of the factors $\hat{\mathbf{z}}_t = E[\mathbf{z}_t | \mathbf{u}_t]$. We then z -scored each factor (that is, each element of $\hat{\mathbf{z}}_t$) separately.

The second step was to estimate the horizontal and vertical velocity of the cursor from the z -scored factors using a Kalman filter:

$$\mathbf{x}_t | \mathbf{x}_{t-1} \sim N(A\mathbf{x}_{t-1} + \mathbf{b}, Q) \quad (3)$$

$$\hat{\mathbf{z}}_t | \mathbf{x}_t \sim N(C\mathbf{x}_t + \mathbf{d}, R) \quad (4)$$

where $\mathbf{x}_t \in \mathbb{R}^{2 \times 1}$ is a vector of horizontal and vertical cursor velocity at time step t . We fitted the parameters A , \mathbf{b} , Q , C , \mathbf{d} and R using maximum likelihood by relating the factors to an estimate of the monkeys' intended velocity during the calibration sessions. At each time point, this intended velocity vector either pointed straight from the current cursor position to the target with a speed equal to the current cursor

speed³⁷ (monkey J, first calibration task) or pointed straight from the centre of the workspace to the target with a constant speed (0.15 m s⁻¹, monkey L and monkey J, second calibration task).

Because spike counts were z-scored before factor analysis, $\mu = 0$. Because factors were z-scored before decoding into cursor velocity, $d = 0$. Because calibration kinematics were centred about the centre of the workspace, $b = 0$.

The decoded velocity that was used to move the cursor at time step t was $\hat{\mathbf{x}}_t = E[\mathbf{x}_t | \hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_t]$. We can express $\hat{\mathbf{x}}_t$ in terms of the decoded velocity at the previous time step $\hat{\mathbf{x}}_{t-1}$ and the current z-scored spike count vector \mathbf{u}_t :

$$\hat{\mathbf{x}}_t = M_1 \hat{\mathbf{x}}_{t-1} + M_2 \mathbf{u}_t \quad (5)$$

$$M_1 = A - KCA \quad (6)$$

$$M_2 = K \Sigma_z \beta \quad (7)$$

$$\beta = \Lambda^T (\Lambda \Lambda^T + \psi)^{-1} \quad (8)$$

As part of the procedure for z-scoring factors, Σ_z is a diagonal matrix where the (p, p) element is the inverse of the standard deviation of the p th factor. K is the steady-state Kalman gain matrix. We z-scored the spike counts and the factors in the intuitive mappings so that the perturbed mappings (which were based on the intuitive mappings) would not require a neural unit to fire outside of its observed spike count range.

Perturbed mappings. The perturbed mappings were modified versions of the intuitive mapping. Within-manifold perturbations altered the relationship between factors and cursor kinematics. The elements of the vector $\hat{\mathbf{z}}_t$ were permuted before being passed into the Kalman filter (red arrows, Fig. 1b). This preserves the relationship between neural units and the intrinsic manifold, but changes the relationship between dimensions of the intrinsic manifold and cursor velocity. Geometrically, this corresponds to re-orienting the control space within the intrinsic manifold.

The following equations describe within-manifold perturbations:

$$\hat{\mathbf{x}}_t = M_1 \hat{\mathbf{x}}_{t-1} + M_{2,WM} \mathbf{u}_t \quad (9)$$

$$M_{2,WM} = K \eta_{WM} \Sigma_z \beta \quad (10)$$

where η_{WM} is a 10×10 permutation matrix defining the within-manifold perturbation (that is, the within-manifold perturbation matrix). Each element of a permutation matrix is either 0 or 1. In each column and in each row of a permutation matrix, one element is 1, and the other elements are 0. In other words, $\eta_{WM} \Sigma_z \beta \mathbf{u}_t$ is a permuted version of $\Sigma_z \beta \mathbf{u}_t$.

Outside-manifold perturbations altered the relationship between neural units and factors. The elements of \mathbf{u}_t were permuted before being passed into the factor-analysis model (blue arrows, Fig. 1b). This preserves the relationship between factors and cursor velocity, but changes the relationship between neural units and factors. Geometrically, this corresponds to re-orienting the control space within the neural space and outside of the intrinsic manifold.

The following equations describe outside-manifold perturbations:

$$\hat{\mathbf{x}}_t = M_1 \hat{\mathbf{x}}_{t-1} + M_{2,OM} \mathbf{u}_t \quad (11)$$

$$M_{2,OM} = K \Sigma_z \beta \eta_{OM} \quad (12)$$

where η_{OM} is a $q \times q$ permutation matrix defining the outside-manifold perturbation (that is, the outside-manifold perturbation matrix). In other words, $\eta_{OM} \mathbf{u}_t$ is a permuted version of \mathbf{u}_t .

Choosing a perturbed mapping. We used data from the first 200 trials (monkey J) or 150 trials (monkey L) of closed-loop control during the baseline blocks to determine the perturbation matrix that we would use for the session. The procedure we used had three steps (detailed below). First, we defined a set of candidate perturbations. Second, we predicted the open-loop cursor velocities for each candidate perturbation. Third, we selected one candidate perturbation. We aimed to choose a perturbation such that the perturbed mapping would not be too difficult for the monkeys to use nor so easy that no learning was needed to achieve proficient performance.

For monkey J, we often alternated perturbation types across consecutive days. For monkey L, we determined which type of perturbation we would use each day before the first experiment. That order was set randomly by a computer. We did this in order to avoid a detectable pattern of perturbation types.

The following describes the first step in choosing a perturbed mapping: defining the candidate perturbations. For within-manifold perturbations, η_{WM} is a 10×10 permutation matrix. The total number possible η_{WM} is 10 factorial (3,628,800). We considered all of these candidate within-manifold perturbations.

For outside-manifold perturbations, η_{OM} is a $q \times q$ permutation matrix, where q is the number of neural units. For a population of 90 neural units, there are 90 factorial ($> 10^{100}$) possible values of η_{OM} . Due to computational constraints, we were unable to consider every possible η_{OM} as a candidate perturbation. We used slightly different procedures to determine the candidate outside-manifold perturbations for the two monkeys.

The procedure we used for monkey J is as follows. We permuted the neural units independently. We chose to permute only the neural units with the largest modulation depths (mean number of units permuted, 39 ± 18). Permuting the units with larger modulation depths impacted the monkey's ability to proficiently control the cursor more than would permuting units with smaller modulation depths. For each session, we randomly chose 6 million η_{OM} that permuted only the specified units. This formed the set of candidate outside-manifold perturbations.

The procedure we used for monkey L is as follows. To motivate it, note that the two perturbation types altered the intuitive mapping control space within a different number of dimensions of the neural space for monkey J. Within-manifold perturbations were confined to ten dimensions of the neural space, but outside-manifold perturbations were confined to N dimensions of the neural space (where N is the number of permuted units, 39 on average). Thus, the dimensionality of the space through which the monkey would have to search to find the perturbed control space was different for the two types of perturbed mappings; it was larger for the outside-manifold perturbations than it was for the within-manifold perturbations. We recognized that this difference may have affected the monkey's ability to learn outside-manifold perturbations. For monkey L, we reduced the size of the search space for the outside-manifold perturbations, thereby equalizing the size of the search space for the two perturbation types. We did this by constraining η_{OM} so that the number of possible η_{OM} was equal to the number of candidate within-manifold perturbations. We then considered all η_{OM} to be candidate outside-manifold perturbations. To construct outside-manifold perturbations, we assigned each neural unit to one of eleven groups. The first ten groups had an equal number of neural units. The eleventh group had the remaining neural units. We specifically put the neural units with the lowest modulation depths in the eleventh group. The 10m (where m is the number of neural units per group) neural units with the highest modulation depths were randomly assigned to the first ten groups. We created outside-manifold perturbations by permuting the first ten groups, keeping all the neural units within a group together. Thus, the number of possible η_{OM} is 10 factorial, all of which were considered as candidate outside-manifold perturbations.

We attempted to keep these groupings as constant as possible across days. On some days, one electrode would become unusable (relative to the previous day) as evident from the threshold crossing waveforms. When this occurred, we kept all of the groupings fixed that did not involve that electrode. If an electrode in one of the first ten groups became unusable, we would substitute it with a neural unit from the eleventh group.

The following describes the second step in choosing a perturbed mapping: estimating the open-loop velocities of each candidate perturbation. The open-loop velocity can be thought of as a coarse approximation to how the cursor would move if the monkey did not learn. The open-loop velocity measurement captures how the neural activity updates the velocity of the cursor from the previous time step, whereas the closed-loop decoder (equation (5)) also includes contributions from the decoded velocity at the previous time step ($M_1 \hat{\mathbf{x}}_{t-1}$) as well as from the neural activity at the current time step ($M_2 \mathbf{u}_t$). To compute the open-loop velocity, we first computed the average z-scored spike counts of every neural unit in the first 200 (monkey J) or 150 (monkey L) trials of the baseline block. We binned the spike counts from 300 ms to 1,300 ms (monkey J) or 1,100 ms (monkey L) after the beginning of each trial, and then averaged the spike counts for all trials to the same target. Together, these comprised 8 spike count vectors (one per target). For each of the spike count vectors, we computed the open-loop velocity for the candidate perturbations:

$$\mathbf{x}_{OL}^i = M_{2,p} \mathbf{u}_B^i \quad (13)$$

where \mathbf{u}_B^i is the mean z-scored spike count vector for the i^{th} target. $M_{2,p}$ is $M_{2,WM}$ for within-manifold perturbations and $M_{2,OM}$ for outside-manifold perturbations.

The following describes the third step in choosing a perturbation: selecting a candidate perturbation. For each candidate perturbation, we compared the open-loop velocities under the perturbed mapping to the open-loop velocities under the intuitive mapping on a per-target basis. We needed the velocities to be dissimilar (to induce learning) but not so different that the animal could not control the cursor. For each target, we measured the angles between the 2D open-loop velocity vectors. We also measured the magnitude of the open-loop velocity for the perturbed mapping. For each session, we defined a range of angles (average minimum of range across sessions: mean \pm s.e.m, $19.7^\circ \pm 7.0^\circ$; average maximum of range across sessions: $44.4^\circ \pm 8.9^\circ$) and a range of velocity magnitudes (average minimum of range across sessions, $0.7 \text{ mm s}^{-1} \pm 0.4 \text{ mm s}^{-1}$; average maximum of range across sessions, $5.5 \text{ mm s}^{-1} \pm 4.0 \text{ mm s}^{-1}$). Note that when the monkey controlled the cursor in

closed-loop (equation (5)), the cursor speeds were much greater than these ranges of open-loop velocities. This is because M_1 was nearly an identity matrix for our experiments. Thus, the term $M_1 \hat{x}_{t-1}$ is expected to be larger than the term $M_2 u_t$. We found all candidate perturbations for which the angles and magnitudes for all targets were within the designated ranges. From the candidate perturbations that remained after applying these criteria, we arbitrarily chose one to use as the perturbation for that session.

Amount of learning. This section corresponds to Fig. 2c. For each session, we computed the amount of learning during perturbation blocks as a single, scalar value that incorporated both changes in success rate (percent of trials for which the peripheral target was acquired successfully) and target acquisition time. We sought to use a metric that captured how much the monkeys' performance improved throughout the perturbation block relative to how much it was impaired at the beginning of the perturbation block. Having a single value for each session allowed us to more easily compare learning across sessions and to relate the amount of learning to a variety of properties of each perturbation (Extended Data Fig. 6). We also analysed each performance criterion individually for each monkey without any normalization (Extended Data Fig. 2). We saw consistent differences in learnability. Thus, our results do not rely on the precise form of our learning metric, but the form we used provides a scalar value as a convenient summary metric.

As success rate and target acquisition time are expressed in different units, we first normalized each metric. We found the mean and standard deviation of the success rates and target acquisition times across all non-overlapping 50-trial bins in the baseline, perturbation and washout blocks for each monkey. We then z-scored the success rates and target acquisition times separately for each monkey. Figure 2c shows normalized performance projected onto veridical units.

For each session, we computed the average z-scored success rate and the average z-scored target acquisition time across all bins in the baseline block.

$$P_B = \begin{bmatrix} s_B \\ a_B \end{bmatrix} \quad (14)$$

where P_B is the performance, s_B is the average normalized success rate and a_B is the average normalized acquisition time during the baseline block (monkey J, 386.9 ± 82.5 trials; monkey L, 292.1 ± 43.5 trials).

We also computed the normalized success rates and acquisition times for all bins in the perturbation blocks.

$$P_P(j) = \begin{bmatrix} s_P(j) \\ a_P(j) \end{bmatrix} \quad (15)$$

where $P_P(j)$ is the performance, $s_P(j)$ is the normalized success rate, and $a_P(j)$ is the average normalized acquisition time during the j th 50-trial bin of the perturbation block.

Empirically, we observed that the monkeys' performance during the perturbation blocks did not exceed the performance during the baseline blocks. Therefore, we define a maximum learning vector (L_{\max}) as a vector that extends from the performance in the first bin with the perturbed mapping to the point corresponding to baseline performance (Fig. 2c).

$$L_{\max} = P_B - P_P(1) \quad (16)$$

The length of this vector is the initial performance impairment because it describes the drop in performance that resulted when we switched from the baseline block to the perturbation block (shown in Fig. 3a and Extended Data Fig. 6a). For each bin (j) within the perturbation blocks, we defined a raw learning vector ($L_{\text{raw}}(j)$). This vector extended from the point corresponding to initial performance during the perturbation block to the point corresponding to performance during each bin.

$$L_{\text{raw}}(j) = P_P(j) - P_P(1) \quad (17)$$

We projected the raw learning vectors onto the maximum learning vector. These were termed the projected learning vectors ($L_{\text{proj}}(j)$).

$$L_{\text{proj}}(j) = \left(L_{\text{raw}}(j) \cdot \frac{L_{\max}}{\|L_{\max}\|} \right) \left(\frac{L_{\max}}{\|L_{\max}\|} \right) \quad (18)$$

The lengths of the projected learning vectors relative to the lengths of the maximum learning vectors define the amount of learning in each 50-trial bin ($L_{\text{bin}}(j)$).

$$L_{\text{bin}}(j) = \frac{\|L_{\text{proj}}(j)\|}{\|L_{\max}\|} \quad (19)$$

An amount of learning of 0 indicates that the monkey did not improve performance, and a value of 1 indicates that the monkey fully improved (up to the level during the baseline block). For each session, we computed the amount of learning for all bins, and we selected the largest one as the amount of learning for that session.

$$L_{\text{session}} = \max_j(L_{\text{bin}}(j)) \quad (20)$$

Figure 2c shows the raw learning vectors for one bin in each of two sessions (thick blue and red lines), along with the projected learning vector (thin red line) and the maximum learning vector (dashed grey line) for one of those sessions.

Principal angles between intuitive and perturbed control spaces. This section corresponds to Fig. 3b and Extended Data Fig. 6b. The control spaces for the intuitive and perturbed BCI mappings in our experiments were spanned by the rows of M_2 for the intuitive mapping, $M_{2,\text{WM}}$ for within-manifold perturbations and $M_{2,\text{OM}}$ for outside-manifold perturbations. Because we z-scored spike counts in advance, the control spaces for each day intersected at the origin of the neural space. The two principal angles³⁸ between the intuitive and perturbed control spaces defined the maximum and minimum angles of separation between the control spaces (Fig. 3b).

Required preferred direction changes. This section corresponds to Fig. 3c and Extended Data Fig. 6c. One way in which learning is manifested is by changes in how individual neurons are tuned to the parameters of the movement, in particular the preferred direction^{7,18}. For each session, we sought to compute the required changes in preferred direction for each neural unit that would lead to proficient control of the cursor under the perturbed mapping. One possibility would be to examine the columns of M_2 and $M_{2,\text{P}}$. Each column can be thought of as representing the pushing direction and pushing magnitude of one unit (that is, the contribution of each neural unit to the velocity of the cursor). We could simply estimate the required change in preferred direction by measuring the change in pushing directions for each unit between the intuitive and perturbed mappings. However, this method is not suitable for the following reason. For outside-manifold perturbations for monkey J, we permuted only a subset of the neural units. As a result, the columns of $M_{2,\text{OM}}$ corresponding to the non-permuted units were the same as in M_2 . By estimating the required change in preferred direction as the difference in directional components of M_2 and $M_{2,\text{OM}}$, we would be implicitly assuming that the monkey is capable of identifying which units we perturbed and changing only their preferred directions, which appears to be difficult to achieve in the timeframe of a few hours⁷. Therefore, we sought a more biologically plausible method of computing the required preferred direction changes.

Using a minimal set of assumptions, we computed the firing rates that each unit should show under one particular learning strategy. Then, we computed the preferred direction of each unit using those firing rates and compared them to the preferred directions during the baseline block. The following were the assumptions used to compute the firing rates:

1. We assumed the monkeys would intend to move the cursor to each target at the same velocity it exhibited under the intuitive mapping. Fitts' Law predicts that movement speed depends on movement amplitude and target size³⁹, and these were always the same in our experiments.

2. The firing rates for the perturbed mapping should be as close as possible to the firing rates we recorded when the monkeys used the intuitive mapping. This keeps the predicted firing rates within a physiological range and implies a plausible exploration strategy in neural space.

We used the following procedure to compute the required preferred direction changes. First, we found the average normalized spike count vector u_B^i across time points (300–1,000 ms after the start of the trial) and all trials to each target (i) during the baseline blocks. We minimized the Euclidian distance between u_B^i and u_P^i , the normalized spike count vector for the perturbed mapping (assumption 2), subject to $M_2 u_B^i = M_{2,\text{P}} u_P^i$ (assumption 1). $M_2 u_B^i$ (the open-loop velocity for the intuitive mapping) is known from the baseline block. For a given perturbed mapping (with $M_{2,\text{P}}$), we sought to find u_P^i that would lead to the same open-loop velocity, which has a closed-form solution:

$$u_P^i = u_B^i + M_{2,\text{P}}^T \left(M_{2,\text{P}} M_{2,\text{P}}^T \right)^{-1} (M_2 - M_{2,\text{P}}) u_B^i \quad (21)$$

For each neural unit (k), we computed its preferred direction $\theta_B(k)$ with the intuitive mapping by fitting a standard cosine tuning model.

$$u_B^i(k) = m_k \cdot \cos(\theta_i - \theta_B(k)) + b_k \quad (22)$$

where $u_B^i(k)$ is the k th element of u_B^i , m_k is the depth of modulation, b_k is the model offset of unit k , and θ_i is the direction of the i th target. We also computed the preferred direction of each unit for the perturbed mapping ($\theta_P(k)$) in the same way. Figure 3c shows histograms of

$$|\theta_P(k) - \theta_B(k)| \quad (23)$$

averaged across all units for each session.

Estimation of intrinsic dimensionality. This section accompanies Fig. 4a–c. During all experiments, we identified a ten-dimensional intrinsic manifold (that is, ten factors). Offline, we confirmed this was a reasonable choice by estimating the intrinsic dimensionality of the data recorded in each calibration block. For each day,

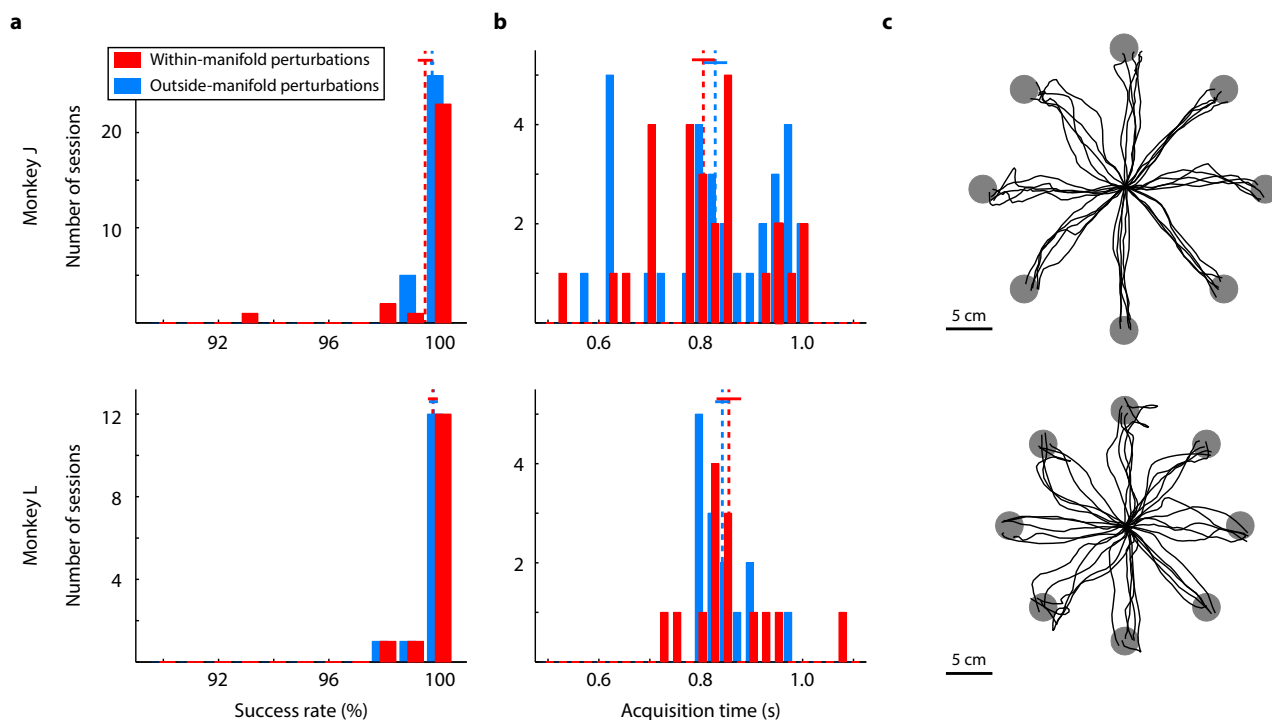
we performed a standard model-selection procedure to compare factor-analysis models with dimensionalities ranging from 2 to 30. For each candidate dimensionality, we used fourfold cross-validation. For each fold, we estimated the factor-analysis model parameters using 75% of the calibration data. We then computed the likelihood of the remaining 25% of the calibration data with the factor-analysis model. For each dimensionality, we averaged the likelihoods across all folds. Each day's 'intrinsic dimensionality' was defined as the dimensionality corresponding to the largest cross-validated data likelihood of the calibration data for that day.

Measuring the cumulative shared variance explained. This section corresponds to Fig. 4d. Factor analysis partitions the sample covariance of the population activity ($\text{cov}(\mathbf{u})$) into a shared component ($\Lambda\Lambda^T$) and an independent component (ψ). In offline analyses, we sought to characterize the amount of shared variance along orthogonal directions within the intrinsic manifold (akin to measuring the lengths of the major and minor axes of an ellipse). These shared variance values are given by the eigenvalues of $\Lambda\Lambda^T$, which can be ordered from largest to smallest. Each eigenvalue corresponds to an 'orthonormalized latent dimension', which refers to identifying orthonormal axes that span the intrinsic manifold. Each orthonormalized dimension is a linear combination of the original ten dimensions. The cumulative shared variance curve is thus informative of how 'oblong' the shared variance is within the manifold, and it can be compared across days. By definition, the cumulative shared variance explained reaches 100% using all ten dimensions, and none of the independent variance (ψ) is explained by those latent dimensions.

Blinding. Investigator blinding was ensured because all sessions were analysed in the same way, by the same computer program. This parallel and automatic treatment of the two perturbation types eliminated investigator biases. The animals were blinded to the test condition delivered each day. If the animals knew which of the two conditions they were presented with, that might have biased our findings. Blinding was achieved before-the-fact with a random and/or unpredictable ordering of experiments, and after-the-fact with control analyses to ensure that conditions were matched as closely as we could detect.

Statistics. For the histograms in Figs 2d and 3, Extended Data Figs 1b, 2 and 7a, the significances of the differences in distributions between within-manifold perturbation samples and outside-manifold perturbation samples were determined with two-tailed Student's *t*-tests assuming unequal variances of the two samples. We ensured that each histogram followed a normal distribution (Kolmogorov–Smirnov test). In Extended Data Figs 1a and 3, the histograms did not follow a normal distribution (Kolmogorov–Smirnov test). For those figures, we used the Wilcoxon rank-sum test to determine the significance of the difference in the distributions. For the linear regressions in Fig. 4 and Extended Data Figs 4 and 6, we determined the significance level of the slopes being different from 0 using *F*-tests for linear regression. We determined whether the difference between two slopes was significant using two-tailed Student's *t*-tests. For all tests, we used $P = 0.05$ as the significance threshold.

31. Tkach, D. C., Reimer, J. & Hatsopoulos, N. G. Observation-based learning for brain-machine Interfaces. *Curr. Opin. Neurobiol.* **18**, 589–594 (2008).
32. Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S. & Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).
33. Santhanam, G. *et al.* Factor-analysis methods for higher-performance neural prostheses. *J. Neurophysiol.* **102**, 1315–1330 (2009).
34. Yu, B. M. *et al.* Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).
35. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. [Ser. A]* **39**, 1–38 (1977).
36. Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P. & Black, M. J. Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Comput.* **18**, 80–118 (2006).
37. Gilja, V. *et al.* A high-performance neural prosthesis enabled by control algorithm design. *Nature Neurosci.* **15**, 1752–1757 (2012).
38. Björck, Å. & Golub, G. H. Numerical methods for computing angles between linear subspaces. *Math. Comput.* **27**, 579–594 (1973).
39. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol. Gen.* **47**, 381–391 (1954).

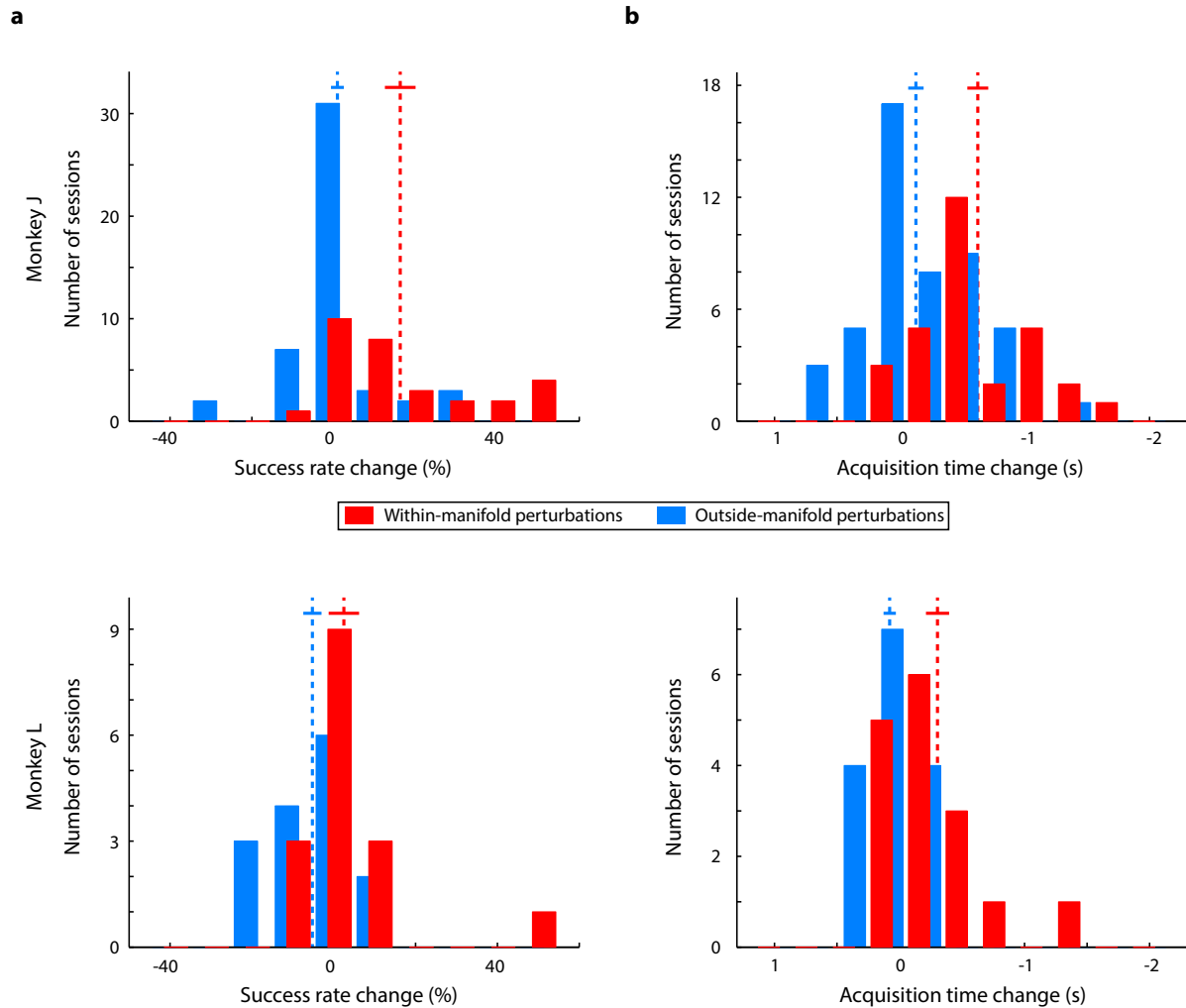


Extended Data Figure 1 | Performance during baseline blocks.

a, Histograms of success rate during the baseline blocks on days when the perturbation would later be within-manifold (red) and outside-manifold (blue) for monkey J (top) and monkey L (bottom). For days with multiple perturbation sessions, the data are coloured according to the first perturbation type. Dashed lines, means of distributions; solid lines, mean \pm s.e.m.

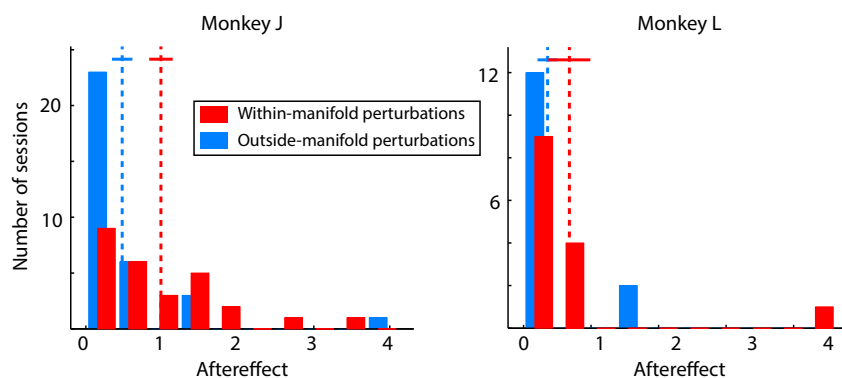
b, Histograms of target acquisition time during baseline blocks. Number of days for panels **a** and **b**: within-manifold perturbations, $n = 27$ (monkey J), 14 (monkey L); outside-manifold perturbations, $n = 31$ (monkey J),

14 (monkey L). **c**, Sample cursor trajectories to all eight targets. At the beginning of each day, the monkeys used the intuitive mapping for 250–400 trials. The monkeys were able to use these mappings to control the cursor proficiently from the outset (as measured by success rate and acquisition time). On all sessions, the success rates were near 100%, and the acquisition times were between 800 and 1,000 ms. No performance metrics during the baseline blocks were significantly different between within-manifold perturbation sessions and outside-manifold perturbation sessions ($P > 0.05$; success rate, Wilcoxon rank-sum test; acquisition time, two-tailed Student's t -test).



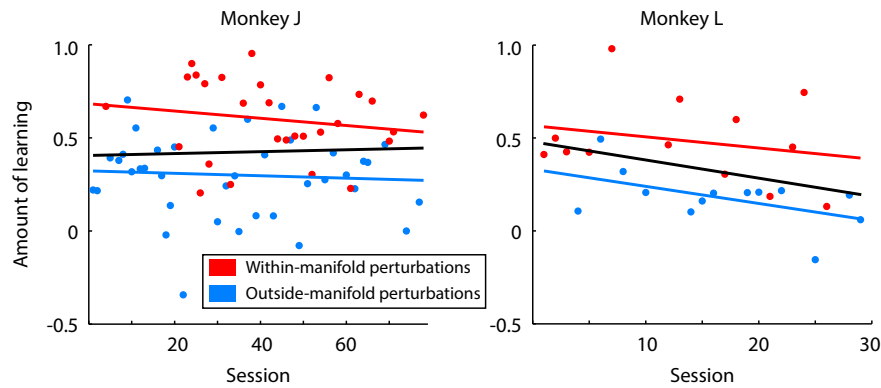
Extended Data Figure 2 | Changes in success rate and acquisition time during perturbation blocks. In Fig. 2d, we quantified the amount of learning in each session using a single metric that combined improvements in success rate and acquisition time. Here, we consider each metric separately. In each comparison, better performance is to the right. **a**, Change in success rate from the first 50-trial bin in the perturbation block to the bin with the best performance. The change in success rate was significantly greater for within-manifold perturbations than for outside-manifold perturbations for monkey J (top, $P < 10^{-3}$, t -test). For monkey L (bottom), the change in success rate was greater for within-manifold perturbations than for outside-manifold

perturbations, and the difference approached significance ($P = 0.088$, t -test). **b**, Change in acquisition time from the first 50-trial bin in the perturbation block to the bin with the best performance. For both monkeys, the change in acquisition time for within-manifold perturbations was significantly greater than for outside-manifold perturbations (monkey J (top), $P < 10^{-4}$, t -test; monkey L (bottom), $P = 0.0014$, t -test). Note that a negative acquisition time change indicates performance improvement (that is, targets were acquired faster). Number of within-manifold perturbations, $n = 28$ (monkey J), 14 (monkey L); outside-manifold perturbations, $n = 39$ (monkey J), 15 (monkey L).



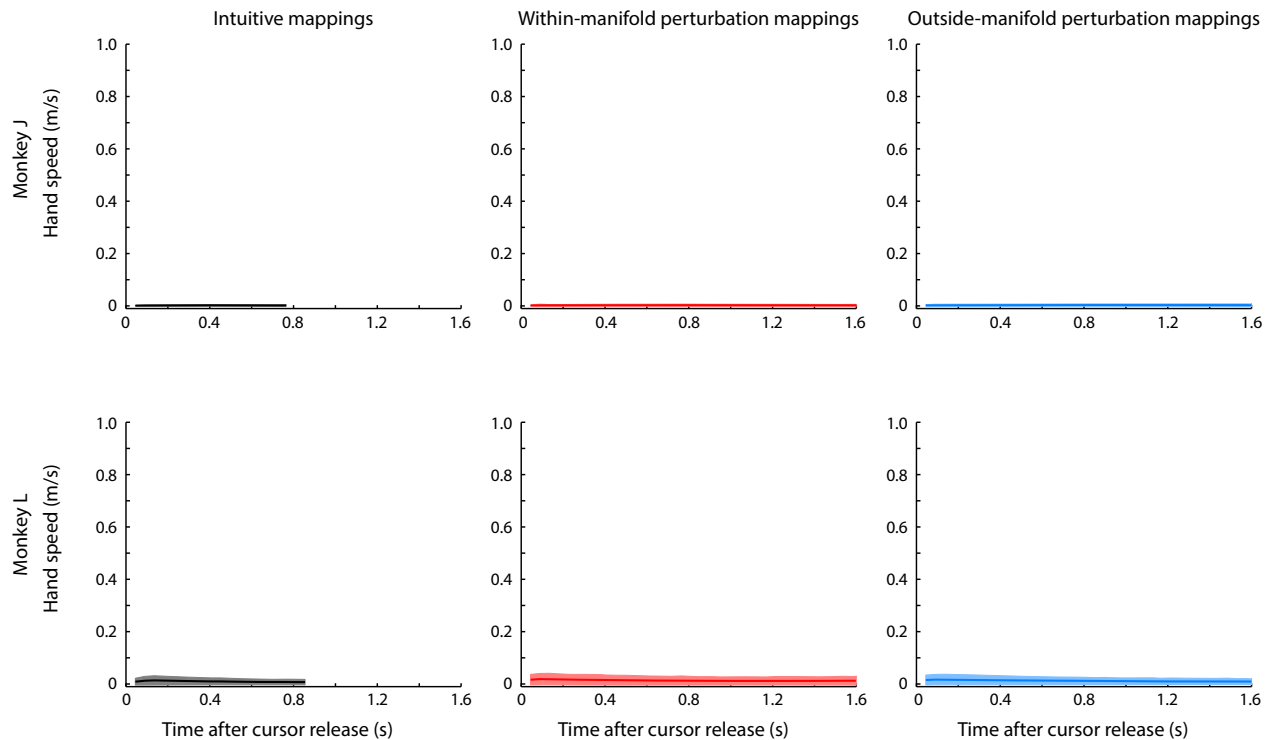
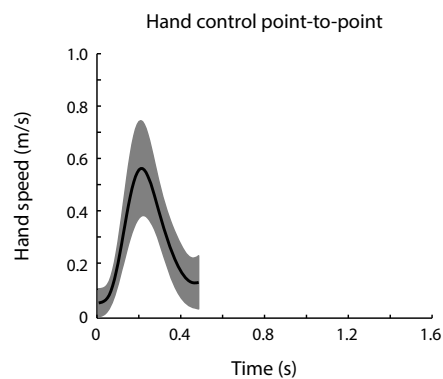
Extended Data Figure 3 | After-effects during washout blocks. After 600 (monkey J) or 400 (monkey L) trials using the perturbed mapping, we re-introduced the intuitive mapping to observe any after-effects of learning. We measured the after-effect as the size of the performance impairment at the beginning of the washout block in the same way that we measured the performance impairment at the beginning of the perturbation block. A larger after-effect indicates more learning had occurred in response to the perturbation. For monkey J (left), the after-effect was significantly larger for

within-manifold perturbations (red) than for outside-manifold perturbations (blue) (Wilcoxon rank-sum test, $P < 10^{-3}$). For monkey L (right), the trend is in the same direction as monkey J, but the effect did not achieve significance (Wilcoxon rank-sum test, $P > 0.05$). These data are consistent with the hypothesis that relatively little learning occurred during the outside-manifold perturbations in comparison to the within-manifold perturbations. Number of within-manifold perturbations, $n = 27$ (monkey J), 14 (monkey L); outside-manifold perturbations, $n = 33$ (monkey J), 15 (monkey L).



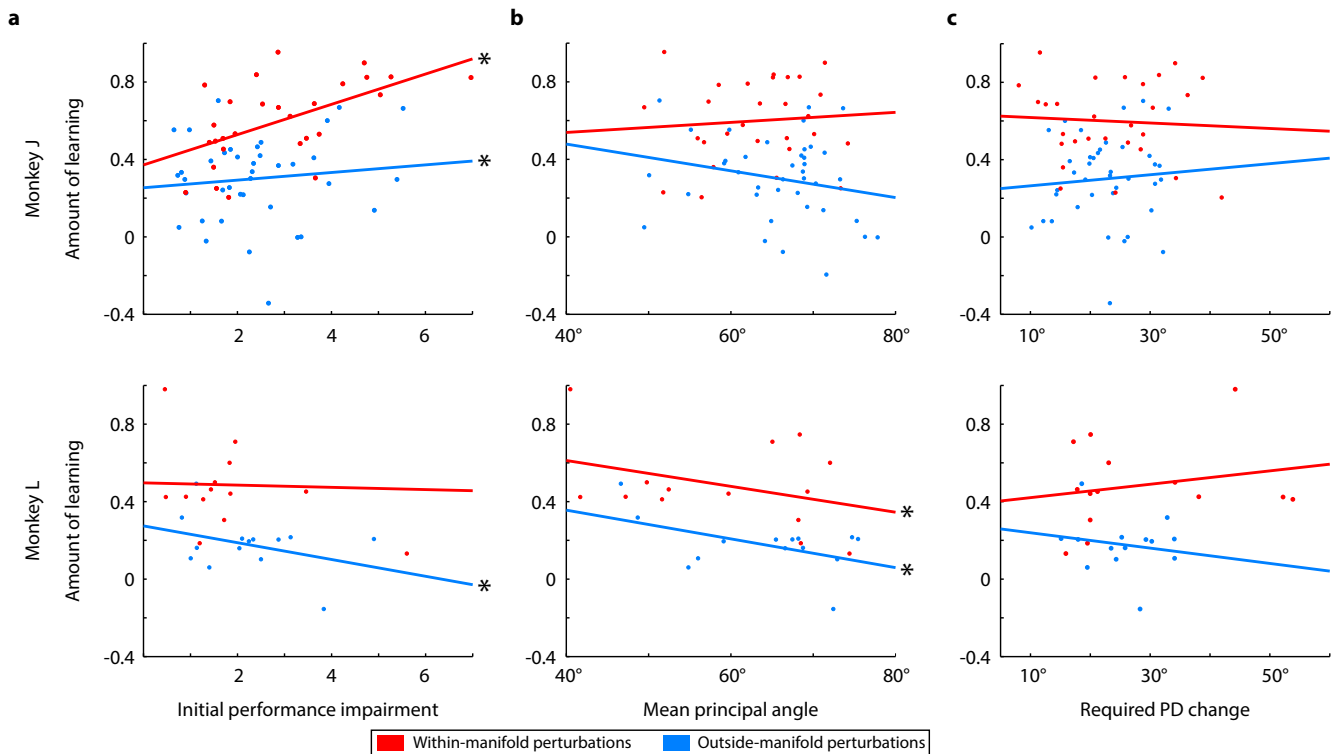
Extended Data Figure 4 | Learning did not improve over sessions. It might have been that, over the course of weeks and months, the animals improved at learning to use perturbed mappings, either one type or both types together. This did not occur. Within-manifold perturbations showed more learning than outside-manifold perturbations across the duration of experiments. Animals

did not get better at learning to use either type of perturbation separately (red and blue regression lines, F -test, $P > 0.05$ for all relationships) nor when considering all sessions together (black regression line, F -test for linear regression, $P > 0.05$). Same number of sessions as in Extended Data Fig. 2. Each point corresponds to one session.

a**b**

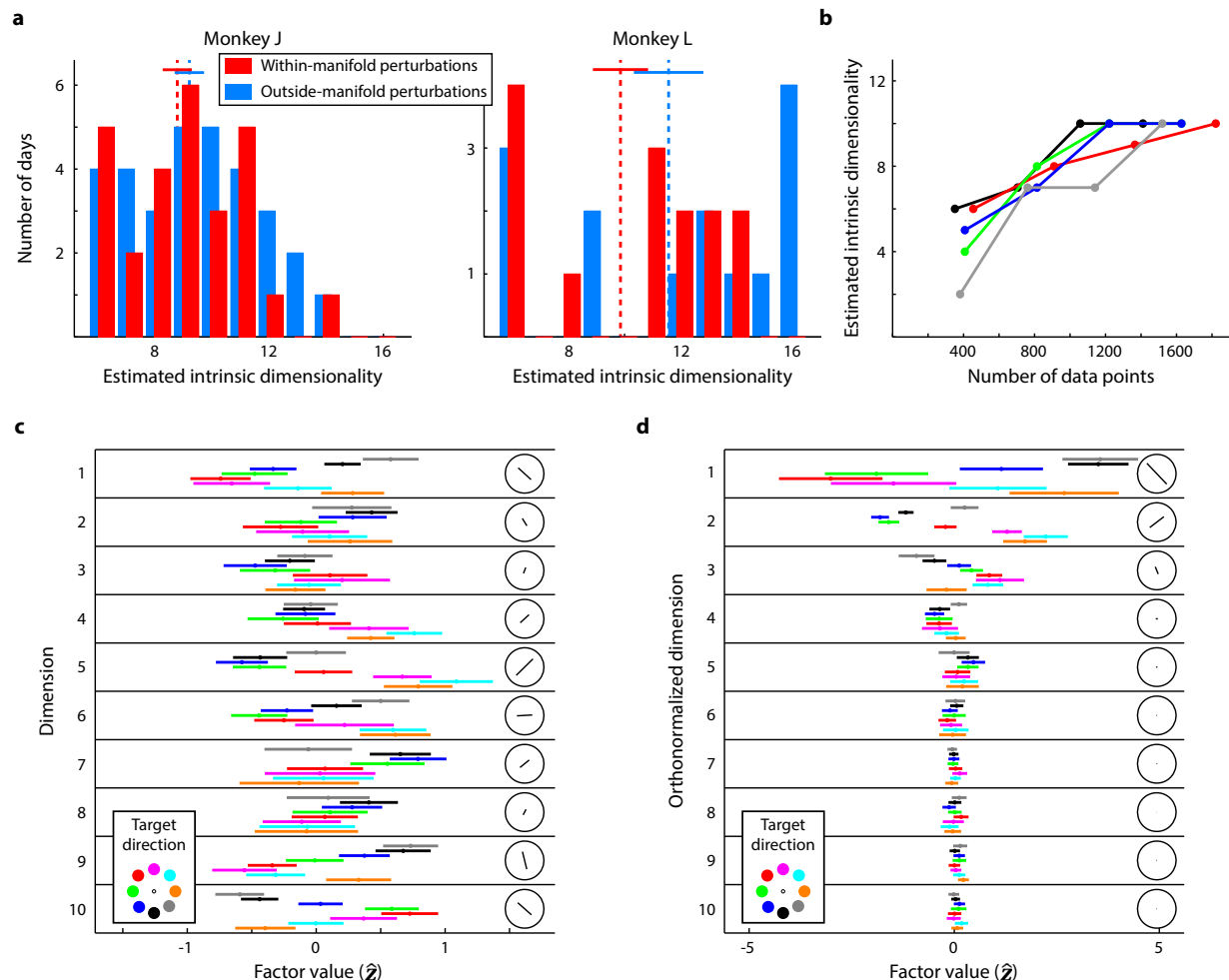
Extended Data Figure 5 | Hand speeds during BCI control and hand control. We loosely restrained the monkeys' arms to the chair's armrests during experiments. The monkeys minimally moved their hands, but the movements did not approach the limits of the restraints. **a**, Average hand speeds across all trials in all sessions for the baseline blocks (left column),

within-manifold perturbation blocks (middle column), and outside-manifold perturbation blocks (right column) for monkey J (top row) and monkey L (bottom row). **b**, Average hand speed during a typical point-to-point reaching task (monkey L). Thus, the hand movements for the BCI tasks are substantially smaller than for the reaching task.



Extended Data Figure 6 | Accounting for within-class differences in learning. **a**, Relation between amount of learning and initial impairment in performance for monkey J (top) and monkey L (bottom). Each point corresponds to one session. Lines are linear regressions for the within-manifold perturbations and outside-manifold perturbations. *Slope significantly different than 0 (F -test for linear regression, $P < 0.05$). **b**, Relation between amount of learning and mean principal angles between control spaces for perturbed and intuitive mappings. **c**, Relation between amount of learning and mean required preferred direction (PD) change. Same number of sessions as in Extended Data Fig. 2. Figure 3 showed that the properties of the perturbed mappings (other than whether their control spaces were within or outside the intrinsic manifold) could not account for differences in learning between the two types of perturbation. However, as is evident in Fig. 2d, within each type of perturbation, there was a range in the amount of learning, including some outside-manifold perturbations that were learnable^{5,7}. In this figure, we examined whether learning within each perturbation type could be accounted for by considering other properties of the perturbed mapping. We regressed the

amount of learning within each perturbation type against the various properties we considered in Fig. 3. Panel **a** shows the initial performance impairment could explain a portion of the variability of learning within both classes of perturbation for monkey J. That monkey showed more learning on sessions when the initial performance impairment was larger. For monkey L, the initial performance impairment could account for a portion of the within-class variation in learning only for outside-manifold perturbations; this monkey showed less learning when the initial performance impairment was larger. We speculate that monkey J was motivated by more difficult perturbations while monkey L could be frustrated by more difficult perturbations. Panel **b** shows that the mean principal angles between control planes were related to learning within each class of perturbation for monkey L only. Larger mean principal angles between the control planes led to less learning. Panel **c** shows that the required PD changes were not related to learning for either type of perturbation for both monkeys. This makes the important point that we were unable to account for the amount of learning by studying each neural unit individually.



Extended Data Figure 7 | Offline analyses of intrinsic manifold properties.

a, The intrinsic dimensionalities for all sessions for monkey J (left) and monkey L (right). For both monkeys, the intrinsic dimensionalities were not significantly different between days when we performed within-manifold perturbations and days when we performed outside-manifold perturbations (t -test, $P > 0.05$). Dashed lines, means of distributions; solid lines, mean \pm s.e.m. Same number of days as in Extended Data Fig. 1. **b**, Relation between intrinsic dimensionality and the number of data points used to compute intrinsic dimensionality. For each of 5 days (one curve per day), we computed the intrinsic dimensionality using 25%, 50%, 75% and 100% of the total number of data points recorded during the calibration block. As the number of data points increased, our estimate of the intrinsic dimensionality increased in a saturating manner. **c**, Tuning of the raw factors. These plots exhibit the factors that were shuffled during within-manifold perturbations. We show for one typical day the average factors (\hat{z}) corresponding to the ten dimensions of the intrinsic manifold over a time interval of 700 ms beginning 300 ms after the start of every trial. Within each row, the coloured bars indicate the mean \pm standard deviation of the factors for each target. The line in each circular inset indicates the axis of 'preferred' and 'null' directions of the factor. The length of the axis indicates the relative depth of modulation. The tuning is along an axis (rather than in a single direction) because the sign of a given factor is arbitrary. **d**, Tuning of the orthonormalized factors. Same session and plotting format as **c**. The orthonormalized dimensions are ordered by the amount of shared variance explained, which can be seen by the variance of the factors across all targets. Note that the axes of greatest variation are separated by approximately 90° for orthonormalized dimensions 1 and 2. This property was typical across days. The retrospective estimate of intrinsic dimensionality (Fig. 4 and Extended Data Fig. 7a) may depend on the richness of the behavioural task, the size of the training set (Extended Data Fig. 7b),

the number of neurons, the dimensionality reduction method and the criterion for assessing dimensionality. Thus, the estimated intrinsic dimensionality should only be interpreted in the context of these choices, rather than in absolute terms. The key to the success of this experiment was capturing the prominent patterns by which the neural units co-modulate. As shown in Fig. 4d, the top several dimensions capture the majority of the shared variance. Thus, we believe that our main results are robust to the precise number of dimensions used during the experiment. Namely, the effects would have been similar as long as we had identified at least a small handful of dimensions. Given the relative simplicity of the BCI and observation tasks, our estimated intrinsic dimensionality is probably an underestimate (that is, a richer task may have revealed a larger set of co-modulation patterns that the circuit is capable of expressing). Even so, our results suggest that the intrinsic manifold estimated in the present study already captures some of the key constraints imposed by the underlying neural circuitry. The probable underestimate of the 'true' intrinsic dimensionality may explain why a few nominal outside-manifold perturbations were readily learnable (Fig. 2d). It is worth noting that improperly estimating the intrinsic dimensionality would only have weakened the main result. If we had overestimated the dimensionality, then some of the ostensible within-manifold perturbations would actually have been outside-manifold perturbations. In this case, the amount of learning would tend to be erroneously low for nominal within-manifold perturbations. If we had underestimated the dimensionality, then some of the ostensible outside-manifold perturbations would actually have been within-manifold perturbations. In this case, the amount of learning would tend to be erroneously high for outside-manifold perturbations. Both types of estimation error would have decreased the measured difference in the amount of learning between within-manifold perturbation and outside-manifold perturbations.

Processing properties of ON and OFF pathways for *Drosophila* motion detection

Rudy Behnia¹, Damon A. Clark^{2,3}, Adam G. Carter⁴, Thomas R. Clandinin³ & Claude Desplan^{1,5}

The algorithms and neural circuits that process spatio-temporal changes in luminance to extract visual motion cues have been the focus of intense research. An influential model, the Hassenstein–Reichardt correlator¹, relies on differential temporal filtering of two spatially separated input channels, delaying one input signal with respect to the other. Motion in a particular direction causes these delayed and non-delayed luminance signals to arrive simultaneously at a subsequent processing step in the brain; these signals are then nonlinearly amplified to produce a direction-selective response. Recent work in *Drosophila* has identified two parallel pathways that selectively respond to either moving light or dark edges^{2,3}. Each of these pathways requires two critical processing steps to be applied to incoming signals: differential delay between the spatial input channels, and distinct processing of brightness increment and decrement signals. Here we demonstrate, using *in vivo* patch-clamp recordings, that four medulla neurons implement these two processing steps. The neurons Mi1 and Tm3 respond selectively to brightness increments, with the response of Mi1 delayed relative to Tm3. Conversely, Tm1 and Tm2 respond selectively to brightness decrements, with the response of Tm1 delayed compared with Tm2. Remarkably, constraining Hassenstein–Reichardt correlator models using these measurements produces outputs consistent with previously measured properties of motion detectors, including temporal frequency tuning and specificity for light versus dark edges. We propose that Mi1 and Tm3 perform critical processing of the delayed and non-delayed input channels of the correlator responsible for the detection of light edges, while Tm1 and Tm2 play analogous roles in the detection of moving dark edges. Our data show that specific medulla neurons possess response properties that allow them to implement the algorithmic steps that precede the correlative operation in the Hassenstein–Reichardt correlator, revealing elements of the long-sought neural substrates of motion detection in the fly.

Behavioural and electrophysiological studies in flies have demonstrated that visual motion responses display the fundamental signatures predicted by the Hassenstein–Reichardt correlator (HRC)^{1,4,5}. In *Drosophila*, photoreceptors R1–R6 are required for motion detection, and synapse onto three lamina monopolar cells L1, L2 and L3 (ref. 6), which provide inputs to distinct motion pathways^{2,3,7–11}. L1 feeds into pathways involved in detecting moving light edges^{2,3} while L2, with contributions from L1 and L3, is involved in detecting moving dark edges^{2,3,10} (Fig. 1b). Deeper in the optic lobe, two direction-selective neuronal types, T4 and T5, are also differentially tuned: T4 cells respond to moving light edges while T5 cells respond to moving dark edges¹². Both T4 and T5 are required for downstream, direction-selective responses of the visual system output cells called lobula plate tangential cells^{13,14}. According to the HRC model, these light and dark edge direction-selective pathways each require two processing steps: differential temporal delay and nonlinear amplification (Fig. 1a). Moreover, these two pathways must process changes in contrast differently to respond differentially to light and dark edges. One such asymmetric mechanism is ‘half-wave rectification’, where inputs of one polarity are amplified and inputs of the opposite polarity are suppressed.

Since L1 and L2 relay information about both contrast increments and decrements³ (they hyperpolarize in response to light increments and depolarize in response to decrements) and act as inputs to motion pathways, we focused on medulla neurons that link L1 and L2 to T4 and T5 to identify potential sites of half-wave rectification and delay. Electron microscopic reconstruction of the medulla has identified columnar neuron types Tm3 and Mi1 as receiving the large majority of synapses

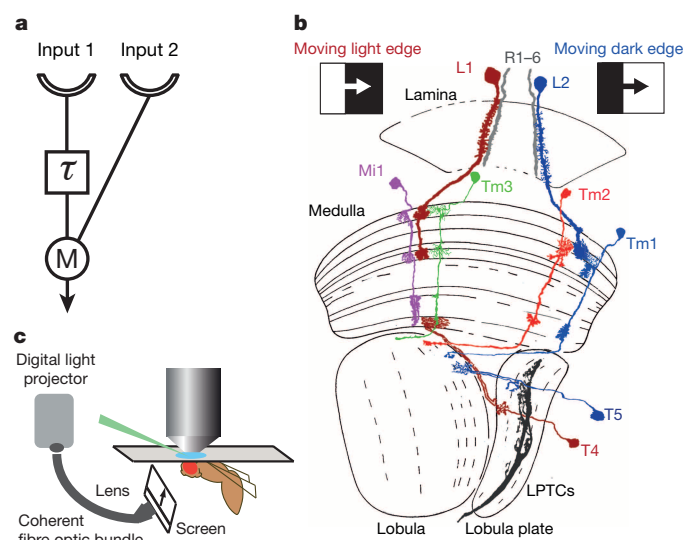


Figure 1 | Motion detection and the fly optic lobe. **a**, A half HRC sensitive to rightward motion. An object moving from left to right first activates input 1 and then input 2. The signal from input 1 is delayed (τ) and arrives at the correlation stage (M for multiplication) close in time to the signal from unit 2, nonlinearly enhancing the signal. For leftward motion, the signals are separated in time by the delay and no motion signal is generated. In the full correlator model, two mirror symmetric correlators are subtracted, producing responses that have opposite signs for opposite directions (see Fig. 4a). **b**, Light edge (L1) and dark edge (L2) motion-sensitive pathways in the *Drosophila* optic lobe. L1 and L2 lamina monopolar cells in the lamina provide inputs to two distinct motion-sensitive pathways that selectively respond to moving light edges and dark edges, respectively. L1 and L3 also contribute to the pathway detecting moving dark edge (not shown). T4 and T5 in the lobula complex are the main inputs to lobula plate tangential cells (LPTCs), and are themselves direction selective. T4 neurons respond selectively to moving light edges and T5 neurons respond to moving dark edges. Mi1 and Tm3 are the main postsynaptic targets of L1 while Tm1 and Tm2 are the main postsynaptic targets of L2. The axons of Mi1 and Tm3 contact T4 in the most proximal medulla layer, whereas Tm1 and Tm2 contact T5 dendrites in superficial lobula layers (modified from ref. 30). **c**, *In vivo* electrophysiology set up: a window is cut in a dorsal region of the head cuticle of an immobilized live fly to expose the cell bodies of medulla neurons to a glass pipette used to perform the recordings. Grey-scale images are displayed on a screen positioned in front of the fly, using a digital light projector coupled to a coherent fibre optic bundle.

¹Center for Developmental Genetics, Department of Biology, New York University, New York, New York 10003-6688, USA. ²Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA. ³Department of Neurobiology, Stanford University, Stanford, California 94305, USA. ⁴Center for Neural Science, New York University, New York, New York 10003, USA. ⁵Center for Genomics & Systems Biology, New York University Abu Dhabi Institute, Abu Dhabi, United Arab Emirates.

from L1 (refs 15, 16). Similarly, Tm1 and Tm2 are the main synaptic targets of L2 (refs 15, 16). Furthermore, Mi1 and Tm3 together constitute over 80% of the presynaptic inputs to T4 (ref. 16) and both Tm1 and Tm2 provide input to T5 (ref. 17). Based on their innervation patterns, Mi1 and Tm3 have been proposed to be core components of the motion detector for light edges involving T4 (ref. 16). Similarly, Tm1 and Tm2 are likely candidates for analogous roles in a motion detector for dark edges.

Since changes in luminance are central to motion detection, we first examined the responses of Mi1, Tm3, Tm1 and Tm2 to step changes in light intensity by performing whole-cell current-clamp recordings on awake immobilized fruit flies (Fig. 1c). Both Mi1 and Tm3 responded with a strong, transient depolarization at the onset of a 1 s light step, and then transiently hyperpolarized to below pre-stimulus levels at light offset (Fig. 2a right and Extended Data Fig. 1a). The responses to onset and offset of light differed in magnitude: in Mi1, the offset hyperpolarization amplitude was only 11% (s.e.m. = 3.5%) of the onset depolarization, while in Tm3 this fraction was 36.6% (s.e.m. = 7.1%) (Extended Data Fig. 1c). A brief flash of light (200 ms) elicited a sharper depolarization in both cells, with the offset hyperpolarization terminating the depolarization phase of the response (Fig. 2a left). The responses observed in Tm1 and Tm2 were similar to each other, yet were strikingly different from those in Mi1 and Tm3 (Fig. 2b and Extended Data Fig. 1b). Tm1 and Tm2 hyperpolarized at light onset, and depolarized strongly at light offset. The hyperpolarization of Tm1 evoked by stimulus onset was 26.1% (s.e.m. = 3.8%) as large as the depolarization evoked at offset; for Tm2, this number was 17.7% (s.e.m. = 2.3%) (Extended Data Fig. 1c). Finally,

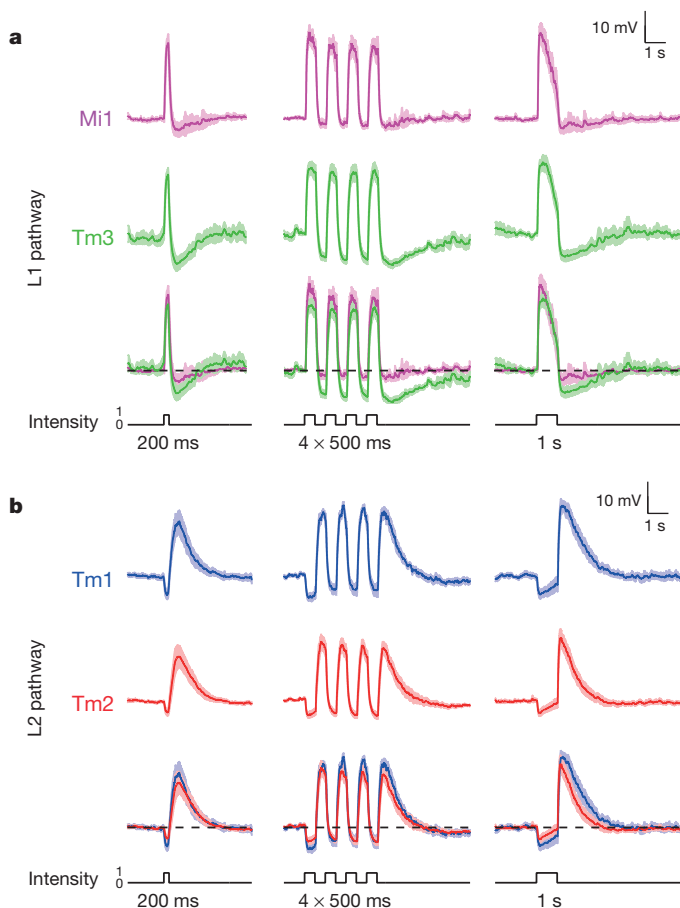


Figure 2 | Mi1/Tm3 respond selectively to brightness increments while Tm1/Tm2 respond selectively to brightness decrements. **a**, Average evoked responses of Mi1 ($n = 7$) and Tm3 ($n = 10$) in the L1 pathway, in response to 200 ms, four consecutive 500 ms and 1 s full-field flashes of light from dark. Thick lines indicate the mean; shaded region indicates s.e.m. **b**, Averaged evoked responses (\pm s.e.m.) of Tm1 ($n = 10$) and Tm2 ($n = 11$) in the L2 pathway in response to the same stimuli.

rapid sequential presentations of light caused repolarization of these cells while their membrane potential was still peaking or decaying from a previous flash (Fig. 2b middle). Thus, Mi1 and Tm3, the postsynaptic targets of L1, respond mostly to brightness increments. Conversely, Tm1 and Tm2, the postsynaptic targets of L2, respond most strongly to brightness decrements, consistent with calcium imaging studies of Tm2 (ref. 18). All four cells showed asymmetries in their responses to brightness changes, consistent with a role in conferring edge selectivity to each pathway. In addition, we examined whether these medulla neurons could relay long-term information about contrast to downstream circuitry by characterizing responses evoked by 5 s brightness increments or decrements presented on an intermediate grey background (Extended Data Fig. 2a–d). All four neurons displayed a sustained response for both brightness increments and decrements, consistent with observations that motion responses can be evoked even when a sequential change in luminance at two points in space occurs with a delay period of up to 10 s in experiments using apparent motion stimuli^{3,19}.

In HRC models, the individual inputs to the cells that perform the nonlinear amplification step are not themselves direction selective. We therefore tested the responses of Mi1, Tm3, Tm1 and Tm2 to motion stimuli, using light and dark bars moving in different directions, under conditions that evoke strong responses from lobula plate tangential cells²⁰ (Extended Data Fig. 3). All four neurons responded to moving bars with a sharp depolarization (Extended Data Fig. 3a, b) but the amplitude of these responses was independent of the direction of motion (Extended Data Fig. 3c). Thus, Mi1, Tm3, Tm1 and Tm2 are not direction selective under these conditions, consistent with these cells acting upstream of the nonlinear correlation stage of motion detection, as recently reported for Tm2 (ref. 18).

We next examined whether Mi1 and Tm3, or Tm1 and Tm2, have different response latencies that would allow them to differentially delay responses to contrast changes. To quantitatively capture the responses of these neurons to dynamic stimuli spanning a wide range of contrast values and time-scales, we used an approximately Gaussian-distributed random flicker stimulus with a 50% contrast (standard deviation) and an exponential correlation time of 10 ms (see Methods). We used standard procedures to extract the linear filter that best described the temporal properties of the response^{21,22} (see Methods). The responses of Mi1 and Tm3 to the noise stimuli were very similar ($r = 0.91$ between mean response traces), with temporal filters that comprised a large positive lobe reflecting a sign-conserving relationship between the contrast input and the neural response (Fig. 3a, b left (arrow) and Extended Data Figs 4a and 5a, b). Mean Tm1 and Tm2 responses to these noise stimuli were also similar to one another ($r = 0.90$), with temporal filters that included a large negative lobe, reflecting a sign-inversion between the contrast input and the neural response (Fig. 3d, e left, arrow; Extended Data Figs 4b and 5e, g). For Mi1 the average peak response time was 71 ms after a contrast change (s.e.m. = 3.8 ms) while it was 53 ms (s.e.m. = 5.2 ms) for Tm3. Thus, a difference of 18 ms existed between the peak times of the filters for Mi1 and Tm3 (Fig. 3b right; $P < 0.01$). Similarly, the average peak time was 56 ms (s.e.m. = 3.8 ms) for Tm1 and 43 ms (s.e.m. = 2.7 ms) for Tm2. The difference in latency between the two cells was 13 ms (Fig. 3e right; $P < 0.002$). Thus, in both cases, there was a small but significant temporal offset, with Mi1 exhibiting a delayed response compared with Tm3, and Tm1 being delayed relative to Tm2. Notably, these peak delay differences are not much smaller than delays inferred from some lobula plate tangential cells recordings and behavioural experiments^{3,23,24}.

We next asked whether neuronal responses to this stochastic stimulus were linear or whether different gains were applied to brightness increments and decrements. The noise stimulus was convolved with the corresponding filters for each neuron type to obtain the predicted linear response of each neuron. We then compared the linear predictions with the actual response to define the instantaneous nonlinearity for each neuron. Consistent with the light step results, the nonlinearities extracted for Mi1 and Tm3 revealed that these cells respond more strongly to brightness increments than to decrements (Fig. 3c and Extended Data Fig. 5b, d).

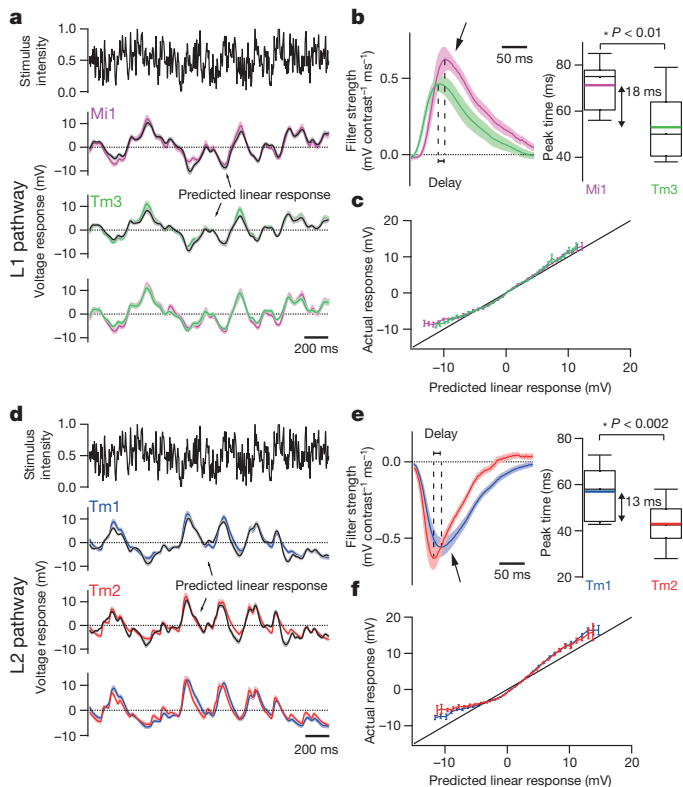


Figure 3 | Mi1/Tm3 and Tm1/Tm2 respond with different delays and nonlinearities to a Gaussian noise stimulus. **a**, Top: 2 s excerpt of the intensity signal from a 10 s full-field Gaussian noise stimulus. Signal correlation time was 10 ms (see Methods). Middle: average voltage response (\pm s.e.m.) of Mi1 ($n = 7$) and Tm3 ($n = 11$) to the 2 s noise stimulus on top. The black trace corresponds to the average predicted linear response (\pm s.e.m.) obtained by convolving the stimulus with the filters in (**b**). Bottom: overlay of the Mi1 and Tm3 responses showing the high similarity in their response. **b**, Left: average linear filters extracted from the data in **a** that best predict the measured response of Mi1 and Tm3 as a function of preceding light intensity changes (\pm s.e.m.). The filters comprise a large positive lobe (arrow) and shallow negative lobe (see Extended Data Fig. 5). Right: box plots of the distribution of the timing of the peak responses of the Mi1 and Tm3 neurons. There is, on average, an 18 ms delay between the peak of Mi1 filters and Tm3 filters. Black line is the median, coloured line is the average. **c**, Average actual responses of Mi1 and Tm3 plotted against their average linear predicted responses. Error bars, s.e.m. A line of slope 1 is shown in black. **d–f**, Same as above for Tm1 ($n = 15$) and Tm2 ($n = 14$). The filters comprise a large negative lobe (arrow) and a shallow positive lobe (Extended Data Fig. 5). The average peak delay between the peak of Tm1 and Tm2 filters is 13 ms.

Similarly, both Tm1 and Tm2 neurons were less hyperpolarized in response to brightness increments and more depolarized in response to brightness decrements than predicted by the linear model (Fig. 3f and Extended Data Fig. 5f, h). The noise stimuli evoked smaller response asymmetries than those observed with brightness steps, possibly because these stimuli use smaller changes in intensity than our step stimuli. Such differences in gain for brightness increments and decrements reflect partial half-wave rectification, a central feature of models that selectively respond to one contrast polarity^{2,3,19,25}.

Can the dynamics of the linear filters and the extent of the nonlinearities we measured account for well-characterized properties of motion detecting pathways? One hallmark of the HRC is that it displays a peak response to a defined temporal frequency, creating a temporal frequency optimum. Because of its structure, the output of an HRC is not proportional to the speed of motion, but rather increases to a maximum value, before decaying at faster speeds. The shape of this tuning curve depends on the temporal properties of its two input channels²⁶. We constructed two separate model correlators, one that used Mi1 and Tm3 filters and

nonlinearities as the two channels preceding multiplication and subtraction, and a second one that used Tm1 and Tm2 filters and nonlinearities (Fig. 4a and data from Extended Data Fig. 6). We examined whether these model motion detectors produced temporal frequency tuning curves similar to those previously measured in flies. When we presented these models with moving sine waves of 20% contrast at various contrast frequencies, we observed a peak response at approximately 1 Hz for both the Mi1/Tm3 and the Tm1/Tm2 models (Fig. 4b). This computed temporal frequency optimum compares favourably with the optima measured in blowflies and *Drosophila*¹⁴.

Another measured feature of these two motion pathways is their selectivity for edges of particular contrast polarity. We presented our model correlators with light and dark edges of 100% contrast, moving across a grey background. The Tm1/Tm2 model was highly selective for dark edges over a range of speeds. The Mi1/Tm3 model was only mildly selective for light edges, owing to the more linear responses measured in Mi1 and Tm3 compared with Tm1 and Tm2 (Fig. 4c). These modelling results are consistent with experimentally measured high selectivity of the dark edge motion pathway^{2,3}, and a more modest selectivity of the light edge motion pathway³.

Taken together, our data are consistent with a model in which Mi1 and Tm3 represent central components of the input channels of a correlator detecting moving light edges, while Tm1 and Tm2 represent analogous components for a correlator that is tuned to detect moving dark edges. The asymmetric responses of these four neurons to brightness increments and decrements corroborates previous studies that argued for separate processing of ON and OFF inputs by distinct channels^{15,25}.

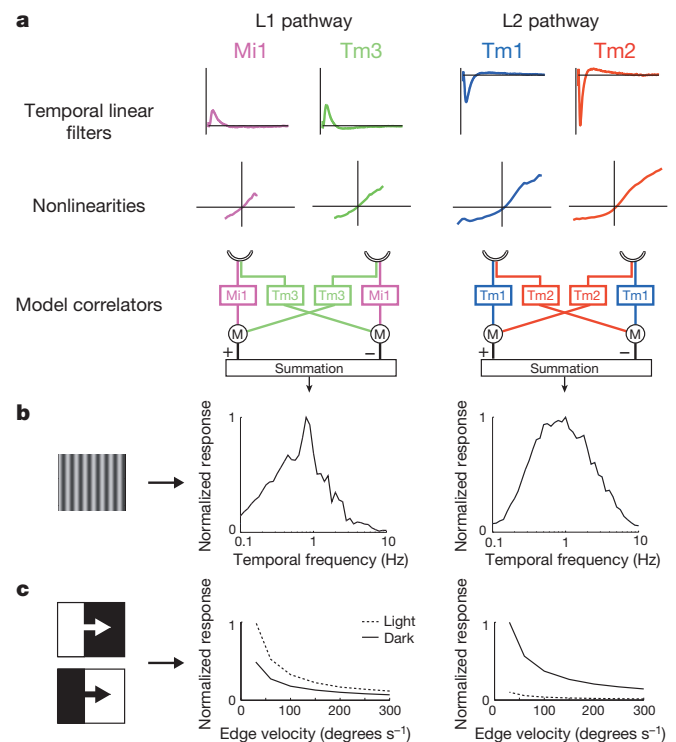


Figure 4 | Modelling Mi1/Tm3 and Tm1/Tm2 as the delayed and non-delayed channels of light edges and dark edges correlators. **a**, Left: the average filters and nonlinearities from the receptive field stochastic data set (200 s noise presentation, 50 ms correlation time, Extended Data Fig. 6) were used to model Mi1 and Tm3 as the delayed and non-delayed channels of a correlator model in bottom left. Right: same as left with Tm1 and Tm2 as the delayed and non-delayed channels of correlator model. **b**, Computed normalized temporal frequency tuning curves obtained numerically for the Mi1/Tm3 model correlator (left) and the Tm1/Tm2 model correlator (right) using sine waves of different temporal frequencies. **c**, Computed normalized response of the Mi1/Tm3 model correlator (left) and the Tm1/Tm2 model correlator (right) to light or dark edges of 100% contrast, moving at a range of speeds.

to explain the segregation between light and dark edge processing. This situation is similar to separate processing of ON and OFF signals by bipolar cells in the vertebrate retina²⁷.

The relative delays we measured between the peak responses in these cells is roughly ten times smaller than previously calculated for idealized motion detector models that fit a host of experimental data. In classic HRC models, input to one channel is not filtered, while input to the second channel is low-pass filtered with a time constant of τ . In these models, the maximum response occurs at a temporal frequency of $1/2\pi\tau$, so that the delay for a 1 Hz optimum is $\tau \sim 150 \text{ ms}$ ²⁶. However, since both filters we measured act as band-pass filters, they suppress high-frequency inputs, while still producing delay differences between the channels. Thus, when input channels contain both these measured filters, a peak timing difference of $\sim 15 \text{ ms}$ can result in a temporal frequency optimum of 1 Hz (see Methods and Extended Data Fig. 7). Furthermore, two considerations might lengthen the actual relative delays between pathways. First, we performed somatic recordings that may only approximate the true axonal response of the neurons. Second, the synapses between Mi1/Tm3 and T4, and those between Tm1/Tm2 and T5 could impose additional delays to either input channel before a correlation operation.

Anatomical reconstruction of the *Drosophila* medulla connected the predicted spatial receptive fields of Mi1 and Tm3 cells to the dendritic arbors of individual T4 cells with known directional preferences¹⁶. According to predictions derived from that analysis, if Mi1 signals are delayed relative to those of Tm3, as our recordings indicate, the observed direction selectivity in T4 could be accomplished by combining Mi1 and Tm3 inputs with opposite signs onto T4 (one inhibitory and the other excitatory). Such an arrangement could be similar to the motion detection model proposed to explain direction selective responses in the vertebrate retina^{16,28}.

Given the cellular and synaptic complexity of medulla circuitry, as well as the wealth of distinct behaviours that are guided by visual motion, additional cell types are likely to play computational roles in *Drosophila* elementary motion detectors. Nonetheless, our data show that Mi1, Tm3, Tm1 and Tm2 possess response properties that are consistent with implementing the algorithmic steps that precede the correlation operations in the motion detecting pathways in *Drosophila*.

METHODS SUMMARY

Green fluorescent protein (GFP)-targeted recordings were performed as described previously²⁹ under visual control using an Olympus microscope and a $\times 60$ objective. Electrophysiology data were collected using Igor (Wavemetrics) running maffPC (courtesy of M. A. Xu-Friedman).

The drivers 686b-Gal4 and Bsh-Gal4 were used to target Mi1 neurons, R13E12-Gal4 for Tm3, 27b-Gal4 for Tm1, otd-Gal4 for Tm2 (see Methods for details about Gal4 driver lines). A cytosolic variant of UAS-GFP (a gift from G. Turner) was expressed under the control of the drivers to visualize the neurons.

Visual stimulation used a digital light projector coupled to a coherent fibre optic bundle to project images on a screen as described previously³.

Data were analysed using custom-written Matlab code. Filter extraction used standard methods²².

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 November 2013; accepted 1 May 2014.

Published online 6 July 2014.

1. Hassenstein, V. & Reichardt, W. System theoretical analysis of time, sequence and sign analysis of the motion perception of the snout-beetle *Chlorophanus*. *Z. Naturforsch. B* **11**, 513–524 (1956).
2. Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. ON and OFF pathways in *Drosophila* motion vision. *Nature* **468**, 300–304 (2010).
3. Clark, D. A., Bursztyn, L., Horowitz, M. A., Schnitzer, M. J. & Clandinin, T. R. Defining the computational structure of the motion detector in *Drosophila*. *Neuron* **70**, 1165–1177 (2011).

4. Borst, A. & Egelhaaf, M. Principles of visual motion detection. *Trends Neurosci.* **12**, 297–306 (1989).
5. Buchner, E. Elementary movement detectors in an insect visual system. *Biol. Cybern.* **24**, 85–101 (1976).
6. Meinertzhagen, I. & O'Neil, S. Synaptic organization of columnar elements in the lamina of the wild type in *Drosophila melanogaster*. *J. Comp. Neurol.* **305**, 232–263 (1991).
7. Rister, J. *et al.* Dissection of the peripheral motion channel in the visual system of *Drosophila melanogaster*. *Neuron* **56**, 155–170 (2007).
8. Katsov, A. Y. & Clandinin, T. R. Motion processing streams in *Drosophila* are behaviorally specialized. *Neuron* **59**, 322–335 (2008).
9. Tuthill, J. C., Nern, A., Holtz, S. L., Rubin, G. M. & Reiser, M. B. Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron* **79**, 128–140 (2013).
10. Sillescu, M. *et al.* Modular use of peripheral input channels tunes motion-detecting circuitry. *Neuron* **79**, 111–127 (2013).
11. Bausenwein, B., Dittrich, A. & Fischbach, K.-F. The optic lobe of *Drosophila melanogaster*. *Cell Tissue Res.* **267**, 17–28 (1992).
12. Maisak, M. S. *et al.* A directional tuning map of *Drosophila* elementary motion detectors. *Nature* **500**, 212–216 (2013).
13. Schnell, B., Raghu, S. V., Nern, A. & Borst, A. Columnar cells necessary for motion responses of wide-field visual interneurons in *Drosophila*. *J. Comp. Physiol. A* **198**, 389–395 (2012).
14. Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. *Annu. Rev. Neurosci.* **33**, 49–70 (2010).
15. Bausenwein, B. & Fischbach, K.-F. Activity labeling patterns in the medulla of *Drosophila melanogaster* caused by motion stimuli. *Cell Tissue Res.* **270**, 25–35 (1992).
16. Takemura, S.-y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500**, 175–181 (2013).
17. Shinomiya, K. *et al.* Candidate neural substrates for off-edge motion detection in *Drosophila*. *Curr. Biol.* **24**, 1–9 (2014).
18. Meier, M. *et al.* Neural circuit components of the *Drosophila* OFF motion vision pathway. *Current Biol.* (2014).
19. Eichner, H., Joesch, M., Schnell, B., Reiff, D. F. & Borst, A. Internal structure of the fly elementary motion detector. *Neuron* **70**, 1155–1164 (2011).
20. Joesch, M., Plett, J., Borst, A. & Reiff, D. F. Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila melanogaster*. *Curr. Biol.* **18**, 368–374 (2008).
21. Chichilnisky, E. J. A simple white noise analysis of neuronal light responses. *Network* **12**, 199–213 (2001).
22. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36**, 909–919 (2002).
23. Harris, R. A., O'Carroll, D. C. & Laughlin, S. B. Contrast gain reduction in fly motion adaptation. *Neuron* **28**, 595–606 (2000).
24. Marmarelis, P. Z. & McCann, G. D. Development and application of white-noise modeling techniques for studies of insect visual nervous system. *Kybernetik* **12**, 74–89 (1973).
25. Franceschini, N., Riehle, A. & Le Nestour, A. in *Facets of Vision* (eds Stavenga, D. G. & Hardie, R. C.) Ch. 17, 360–390 (Springer, 1989).
26. Borst, A. & Bahde, S. What kind of movement detector is triggering the landing response of the housefly? *Biol. Cybern.* **55**, 59–69 (1986).
27. Wässle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* **5**, 747–757 (2004).
28. Barlow, H. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178**, 477–504 (1965).
29. Wilson, R. I., Turner, G. C. & Laurent, G. Transformation of olfactory representations in the *Drosophila* antennal lobe. *Science* **303**, 366–370 (2004).
30. Fischbach, K.-F. & Dittrich, A. The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258**, 441–475 (1989).

Acknowledgements We thank G. Turner for his initial help on whole-cell recordings; X. Li for providing the confocal image of the 686-Gal4 flip-out clone, and N. Vogt, J. Rister and members of the Clandinin laboratory, as well as H. S. Seung and S. A. Baccus, for reading of the manuscript. We also thank D. B. Chklovskii for suggesting recording from Tm3 and T. Ercllick for identifying the Tm3 Gal4 line. This work was supported by a grant from the National Institutes of Health (NIH) (R01EY017916) and a grant from New York University Abu Dhabi Institute (G1205C) to C.D.; R.B. was supported by fellowships from EMBO and the Human Frontier Science Program. This work was also supported by grants from the NIH to T.R.C. (R01EY022638 and DP1 OD003530). D.A.C. was partly supported by an NIH T32 Vision Training Grant and a Jane Coffin Childs postdoctoral fellowship. A.G.C. was supported by a Scholar Award from The McKnight Foundation.

Author Contributions R.B. designed experiments, performed electrophysiological recordings and analysed the data. D.A.C. designed visual stimuli and experiments, analysed the data and performed modelling. A.G.C. provided electrophysiological training and advice to R.B. T.R.C. and C.D. contributed to the design of experiments. R.B., D.A.C., T.R.C. and C.D. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (rb141@nyu.edu) or D.A.C. (damon.clark@yale.edu).

miR-34a blocks osteoporosis and bone metastasis by inhibiting osteoclastogenesis and Tgif2

Jing Y. Krzeszinski¹, Wei Wei¹, HoangDinh Huynh¹, Zixue Jin¹, Xunde Wang¹, Tsung-Cheng Chang², Xian-Jin Xie^{3,4}, Lin He⁵, Lingegowda S. Mangala^{6,7}, Gabriel Lopez-Berestein^{7,8}, Anil K. Sood^{6,7,9}, Joshua T. Mendell^{2,3} & Yihong Wan^{1,3}

Bone-resorbing osteoclasts significantly contribute to osteoporosis and bone metastases of cancer^{1–3}. MicroRNAs play important roles in physiology and disease^{4,5}, and present tremendous therapeutic potential⁶. Nonetheless, how microRNAs regulate skeletal biology is underexplored. Here we identify miR-34a as a novel and critical suppressor of osteoclastogenesis, bone resorption and the bone metastatic niche. miR-34a is downregulated during osteoclast differentiation. Osteoclastic miR-34a-overexpressing transgenic mice exhibit lower bone resorption and higher bone mass. Conversely, miR-34a knockout and heterozygous mice exhibit elevated bone resorption and reduced bone mass. Consequently, ovariectomy-induced osteoporosis, as well as bone metastasis of breast and skin cancers, are diminished in osteoclastic miR-34a transgenic mice, and can be effectively attenuated by miR-34a nanoparticle treatment. Mechanistically, we identify transforming growth factor- β -induced factor 2 (Tgif2) as an essential direct miR-34a target that is pro-osteoclastogenic. Tgif2 deletion reduces bone resorption and abolishes miR-34a regulation. Together, using mouse genetic, pharmacological and disease models, we reveal miR-34a as a key osteoclast suppressor and a potential therapeutic strategy to confer skeletal protection and ameliorate bone metastasis of cancers.

We examined the levels of several cancer-related microRNAs (miRNAs) during a time course of bone marrow osteoclastogenesis assay (Fig. 1a). While the expression of an osteoclast marker tartrate-resistant acid phosphatase (TRAP) was rapidly increased by RANKL and further elevated by rosiglitazone^{7,8} (Fig. 1b), miR-34a was rapidly downregulated by RANKL and further diminished by rosiglitazone (Fig. 1c). The levels of miR-34b/c, two other members in the miR-34 family, were unaffected and expressed at much lower levels than miR-34a (Fig. 1d).

The sequence of miR-34a is evolutionarily conserved and identical in mice and humans. Osteoclast differentiation both from mouse bone-marrow precursors (Fig. 1e, f) and from human peripheral blood mononuclear cells (hPBMC) (Fig. 1g–j) was inhibited by a miR-34a precursor (pre-miR-34a) but enhanced by an antisense miR-34a inhibitor (anti-miR-34a), indicating that miR-34a regulation of bone resorption in mice will probably translate to human pathophysiology.

We generated osteoclastic miR-34a transgenic mice using CAG34a mice (Fig. 2a) and Tie2-cre mice^{7,8}. Fluorescence-activated cell sorting (FACS) and imaging showed that osteoclast progenitors from the 34a-Tie2-Tg (CAG34a⁺Cre⁺) mice were converted to GFP⁺LacZ⁺ whereas the controls (CAG34a⁺Cre⁻) remained GFP⁻LacZ⁺ (Extended Data Fig. 1a, b). Northern blot confirmed the overexpression of mature-miR-34a in the bone marrow of 34a-Tie2-Tg mice (Extended Data Fig. 1c).

Osteoclast differentiation assay reveals that the higher levels of mature miR-34a in the 34a-Tie2-Tg cultures resulted in a lower induction of osteoclast markers, diminished number/size of mature osteoclasts and reduced resorptive activity, whereas precursor proliferation or survival

was unaltered (Fig. 2b and Extended Data Fig. 1d–g). Consequently, serum bone resorption marker CTX-1 (carboxy-terminal telopeptides of type I collagen) and osteoclast number were decreased, whereas osteoblast number, bone formation rate and mineral apposition rate were unaltered (Fig. 2c and Extended Data Fig. 1h, i).

Micro-computed tomography of the proximal tibiae showed that 34a-Tie2-Tg mice had increased bone mass and decreased structure model index (SMI), which quantifies the relative amount of plates (SMI = 0, strong) and rods (SMI = 3, fragile) (Fig. 2d, e). Cortical BV/TV was also higher (Fig. 2f). Moreover, miR-34a transgenic mice generated by three other osteoclast-targeting cre drivers also exhibited a similar phenotype (Extended Data Figs 2 and 3). Thus, miR-34a in the osteoclast lineage augments bone mass by suppressing osteoclastogenesis and bone resorption.

To determine whether miR-34a is a physiologically relevant regulator of bone resorption, we next examined miR-34a knockout (34a-KO) and heterozygous (34a-Het) mice (Fig. 2g). Northern blot confirmed the diminished levels of miR-34a in 34a-KO (Extended Data Fig. 4a). Consistent with prior reports^{9,10}, miR-34a deletion had no overt effect on mouse development. Osteoclast differentiation was augmented in 34a-Het and 34a-KO cultures, whereas precursor proliferation or survival was unaffected (Fig. 2h and Extended Data Fig. 4b–e). As a result, serum CTX-1 and osteoclast number were elevated (Fig. 2i and Extended Data Fig. 4g, h). Micro-computed tomography revealed that 34a-KO and 34a-Het mice exhibited a low bone mass with decreased connectivity density and increased SMI (Fig. 2j–l). Global miR-34a deletion also decreased bone formation as the serum marker PINP (amino-terminal propeptide of type I procollagen), osteoblast number, bone formation rate and mineral apposition rate were reduced (Fig. 2m and Extended Data Fig. 4g, h). The increased resorption in 34a-Het indicates that miR-34a function is haploinsufficient and sensitive to dosage reduction. The recently published miR-34abc triple knockout (34abc-TKO)¹⁰ and full miR-34a KO⁹ also showed a similar phenotype (Extended Data Fig. 5a–h), which validates our miR-34a gene trap mice and strengthens the finding that miR-34a loss-of-function elevates bone resorption.

Bone marrow transplantation showed that wild-type (WT) mice receiving 34a-KO marrow also exhibited higher CTX-1 (Extended Data Fig. 4f) compared with WT mice receiving WT marrow. Furthermore, osteoclastic miR-34a conditional knockout mice (34a-Tie2-KO) also exhibited elevated osteoclast differentiation and bone resorption, but unaltered bone formation, leading to a decreased bone mass (Extended Data Fig. 5i–n). Thus, miR-34a deletion in the osteoclast lineage elevates bone resorption.

Our genetic findings prompted us to investigate whether pharmacological administration of a miR-34a mimic can attenuate postmenopausal osteoporosis using an ovariectomy (OVX) mouse model and a chitosan (CH) nanoparticle vehicle. Reduction of uterine weight in all ovariectomized mice indicated effective oestrogen depletion (Fig. 3a). Unaltered body weight indicated the absence of obvious toxicity from CH nanoparticles

¹Department of Pharmacology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ²Department of Molecular Biology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ³Simmons Cancer Center, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁴Department of Clinical Sciences, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁵Division of Cellular and Developmental Biology, Molecular and Cell Biology Department, University of California at Berkeley, Berkeley, California 94705, USA. ⁶Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁷Center for RNA Interference and Non-coding RNA, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁸Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁹Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.

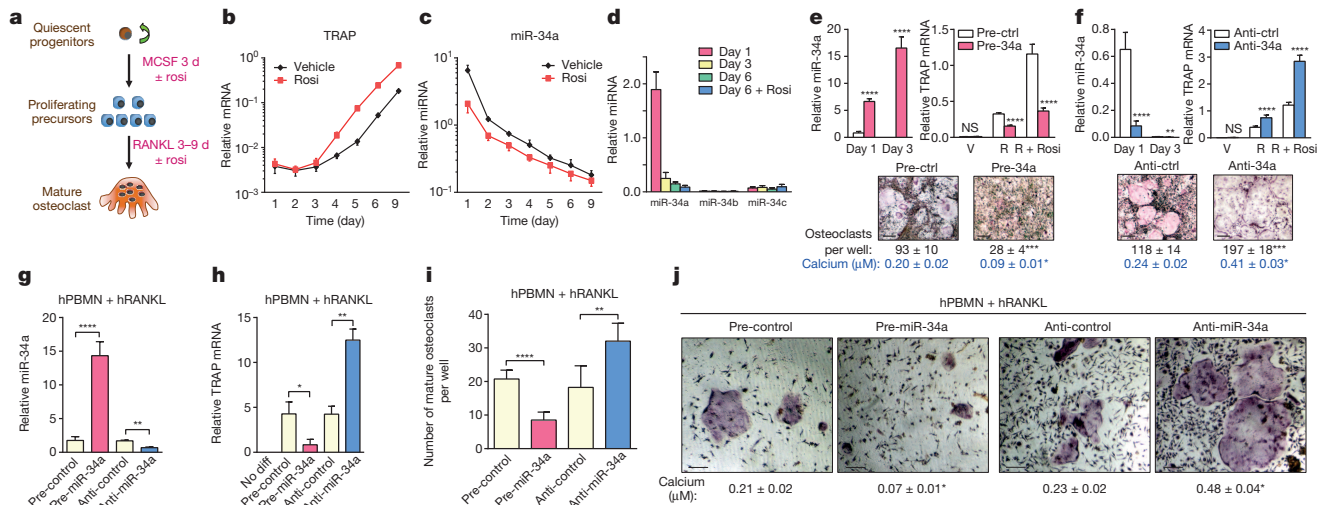


Figure 1 | miR-34a suppresses osteoclastogenesis *ex vivo*. **a**, Diagram of bone marrow osteoclast differentiation assay; rosi, rosiglitazone. **b–d**, TRAP expression (**b**) and mature miRNA levels (**c**, **d**) ($n = 3$). **e**, **f**, Osteoclast differentiation was decreased by pre-miR-34a (**e**) but increased by anti-miR-34a (**f**) ($n = 3$). Left, mature miR-34a levels; right, TRAP expression; bottom, images of TRAP-stained cultures; mature osteoclast numbers (black) and

resorptive activity (blue). Scale bar, 25 μm. V, vehicle; R, RANKL.

g–j, Human RANKL-mediated osteoclast differentiation from hPBMMN cells ($n = 4$). **g**, Mature miR-34a levels. **h**, TRAP expression. **i**, Mature osteoclast numbers. **j**, TRAP staining and resorptive activity. Scale bar, 25 μm.

Error bars, s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$; NS, non-significant.

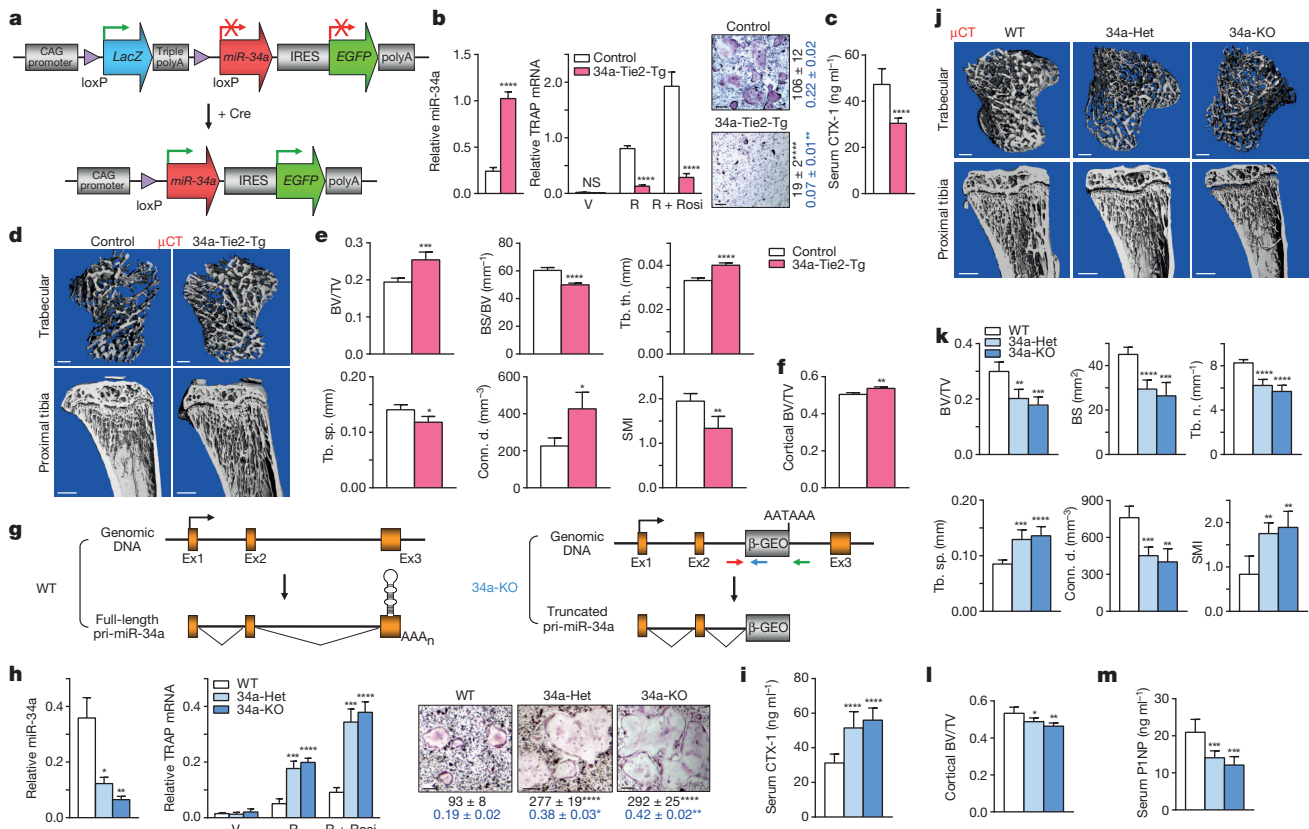


Figure 2 | miR-34a inhibits bone resorption and increases bone mass *in vivo*. **a**, Diagram of the conditional miR-34a transgene (CAG-34a). **b**, 34a-Tie2-Tg cultures showed decreased osteoclast differentiation ($n = 3$). Left, miR-34a levels; middle, TRAP expression; right, TRAP staining, osteoclast numbers (black) and resorptive activity (blue). Scale bar, 25 μm. **c**, Serum CTX-1 (2-month-old, male, $n = 5$). **d–f**, Micro-computed tomography (μCT) of the tibiae (2-month-old, male, $n = 4$). **d**, Images of the trabecular bone of the tibial metaphysis (top) (scale bar, 10 μm) and the entire proximal tibia (bottom) (scale bar, 1 mm). **e**, Trabecular bone parameters. BV/TV, bone

volume/tissue volume ratio; BS/BV, bone surface/bone volume ratio; Tb. th., trabecular thickness; Tb. sp., trabecular separation; Conn. d., connectivity density. **f**, Cortical BV/TV. **g**, Diagram of miR-34a gene-trap knockout. **h**, 34a-KO and 34a-Het cultures showed enhanced osteoclast differentiation ($n = 3$). **i**, Serum CTX-1 (2-month-old, male, $n = 6$). **j–l**, Micro-computed tomography of the tibiae (2-month-old, male, $n = 4$). **j**, Images. **k**, Trabecular bone parameters. Tb. n., trabecular number. **l**, Cortical BV/TV. **m**, Serum P1NP (2-month-old males, $n = 6$). Error bars, s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$; NS, non-significant.

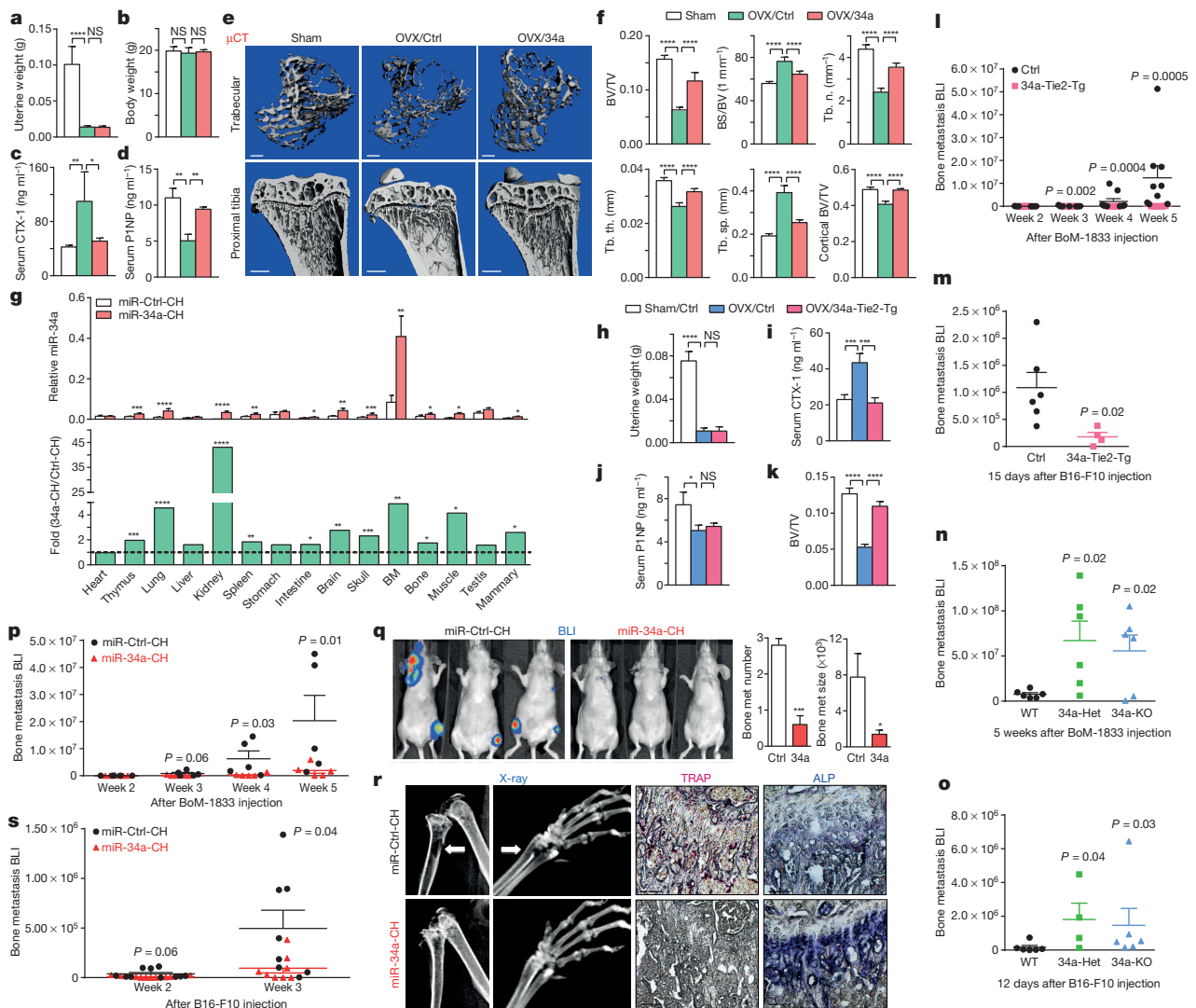


Figure 3 | miR-34a attenuates osteoporosis and cancer bone metastases. **a–f**, OVX or sham operation was performed on 10-week-old female mice. Three days after surgery, the OVX mice were treated with miR-34a-CH (34a) or miR-Ctrl-CH (Ctrl) at 5 μ g per mouse twice a week for 5 weeks ($n = 5$).

a, Uterine weight. **b**, Body weight. **c**, Serum CTX-1. **d**, Serum P1NP. **e**, Micro-computed tomography images. **f**, Trabecular bone parameters. **g**, miR-34a levels in each tissue from miR-34a-CH- vs. miR-Ctrl-CH-treated mice 72 h after a single injection ($n = 3$). Top, mature miR-34a levels; bottom, fold induction. BM, bone marrow. **h–k**, 34a-Tie2-Tg mice or controls (3-month-old, female, $n = 7$) were subjected to OVX and analysed 5 weeks after surgery. **h**, Uterine weight. **i**, Serum CTX-1. **j**, Serum P1NP. **k**, BV/TV by micro-computed tomography. **l**, Xenograft of MDA231-BoM-1833 cells into 34a-Tie2-Tg ($n = 8$) or control ($n = 9$). **m**, Allograft of B16-F10 cells into

34a-Tie2-Tg ($n = 4$) or control ($n = 6$). **n**, MDA231-BoM-1833 cells in 34a-KO ($n = 6$), 34a-Het ($n = 6$) or control ($n = 6$). **o**, B16-F10 in 34a-KO ($n = 6$), 34a-Het ($n = 4$) or control ($n = 6$). **p–r**, Bone metastasis of MDA231-BoM-1833 cells was attenuated by miR-34a-CH delivered 3 days after xenograft at 10 μ g per mouse twice a week for 5 weeks ($n = 5$). **p**, Bioluminescence imaging (BLI) signal. **q**, Left, BLI images; right, number and size of bone metastases (met). **r**, X-ray images and histology images for TRAP and ALP (alkaline phosphatase) staining. Arrows, osteolytic lesions. **s**, Bone metastasis of B16-F10 cells was attenuated by miR-34a-CH delivered at 5 μ g per mouse twice a week for 4 weeks starting 1 week before cancer cell injection ($n = 8$). **l–s**, Statistical analyses used a Mann–Whitney U test and are shown as mean \pm s.d. with P values illustrated. **a**, **c**, **d**, **f**, **h–l**, **n–p**, $P < 0.05$ by analysis of variance. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$; NS, non-significant.

(Fig. 3b). Compared with sham controls, OVX mice treated with miR-Ctrl-CH showed increased CTX-1 and decreased P1NP, whereas both effects were largely prevented in OVX mice treated with miR-34a-CH (Fig. 3c, d). Consequently, OVX-induced bone loss was attenuated by miR-34a-CH (Fig. 4e, f and Extended Data Fig. 6a). miR-34a-CH also decreased bone resorption and increased bone formation in sham controls, leading to a higher bone mass (Extended Data Fig. 6b–d). Biodistribution analysis showed that miR-34a level in the bone marrow was the highest, and further increased fivefold by miR-34a-CH, indicating an efficient miR-34a delivery (Fig. 3g).

In addition to acute systemic miR-34a treatment, we also examined the effects of chronic osteoclastic miR-34a overexpression. OVX-induced bone resorption and bone loss were also attenuated in 34a-Tie2-Tg mice without altering OVX effects on bone formation (Fig. 3h–k and Extended

Data Fig. 6e). These results indicate that osteoclastic miR-34a overexpression is sufficient to impede osteoporosis, and the osteoclast is a key site for miR-34a therapeutic benefit.

To determine whether osteoclastic miR-34a confers protection from bone metastases, we employed two cancer-cell-cardiac-injection models. First, a human breast cancer cell line (MDA231-BoM-1833) was xenografted into female nude mice. This model allowed us to assess cancer cells from human. Second, a mouse melanoma cell line (B16-F10) was allografted into immunocompetent male mice. This model took consideration of adaptive immunity. In both models, bone metastases were attenuated in 34a-Tie2-Tg and 34a-PT-Tg mice but exacerbated in 34a-KO and 34a-Het mice (Fig. 3l–o and Extended Data Fig. 7a–f). Because miR-34a remained intact in the exogenous cancer cells, the altered bone metastases resulted from the altered miR-34a in the bone microenvironment of the host.

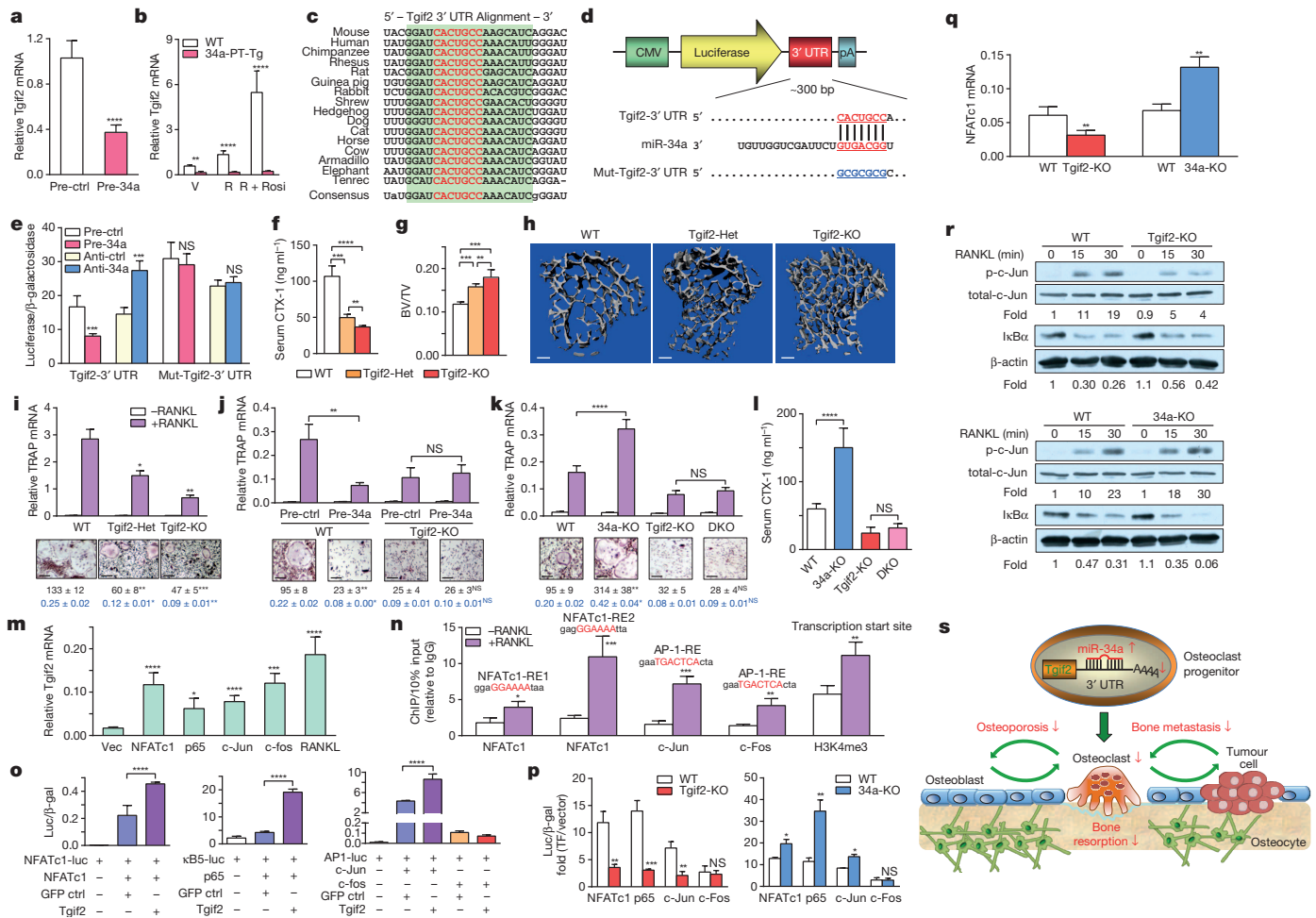


Figure 4 | *Tgfr2* is an essential miR-34a direct target and a pro-osteoclastogenic factor. **a**, *Tgfr2* expression was inhibited by pre-miR-34a in osteoclast cultures (n = 3). **b**, *Tgfr2* expression in WT and 34a-PT-Tg osteoclast cultures (n = 3). **c**, Sequence alignment of the *Tgfr2* 3' UTR. **d**, A diagram of *Tgfr2* 3' UTR reporters. **e**, Luciferase readout from WT or mutant *Tgfr2* 3' UTR reporter co-transfected in HEK293 cells with pre-miR-34a or anti-miR-34a (n = 3). **f-h**, Comparison of *Tgfr2*-KO, *Tgfr2*-Het and WT control mice (1.5-month-old, male, n = 7). **f**, Serum CTX-1. **g-h**, Micro-computed tomography of tibiae. **g**, Trabecular BV/TV. **h**, Images of the trabecular bone of the tibial metaphysis (scale bar, 10 μ m). **i**, Decreased osteoclast differentiation in *Tgfr2*-KO and *Tgfr2*-Het cultures (n = 3). **j**, *Tgfr2*-KO cultures were resistant to the anti-osteoclastogenic effects of pre-miR-34a (n = 3). **i-j**, Top, TRAP expression; bottom, TRAP staining, osteoclast number (black) and resorptive activity (blue). **k-l**, *Tgfr2* 34a double knockout (DKO) mice were compared

Pharmacologically, we tested both a treatment protocol using the human breast cancer model and a prevention protocol using the mouse melanoma model. In both cases, bone metastases were diminished by miR-34a-CH (Fig. 3p-s). Systemic miR-34a-CH delivery affected neither tumour growth nor metastasis to other organs such as lung (Extended Data Fig. 8a, b). Moreover, treating only the cancer cells with miR-34a-CH before injection had no effect (Extended Data Fig. 8c, d). Consistent with the published finding that miR-34abc deletion does not increase tumorigenesis¹⁰, our 34a-KO mice also showed unaltered cancer susceptibility (Extended Data Fig. 8e).

Since systemic miR-34a-CH treatment not only decreases bone resorption but also increases bone formation, we examined the effects of miR-34a overexpression in osteoblasts. We bred the CAG34a mice with Osterix-CreER mice to generate 34a-Osx-Tg mice. Osteoblast differentiation was reduced for 34a-KO and 34a-Het mice, but increased for 34a-Osx-Tg mice (Extended Data Fig. 9a-d). Consequently, 34a-Osx-Tg mice exhibited a higher bone formation but unaltered bone resorption, leading to

an increased bone mass (Extended Data Fig. 9e-g). Importantly, however, the elevated bone formation alone in the 34a-Osx-Tg mice was insufficient to attenuate either OVX-induced bone loss or cancer bone metastases (Extended Data Fig. 9h, i). Together, our conditional miR-34a transgenic mouse models pinpointed the mechanisms underlying the therapeutic benefits of miR-34a by revealing that osteoclast, rather than cancer cell or osteoblast, is the critical and essential player.

To elucidate the mechanisms, we identified *Tgfr2* as a novel direct miR-34a target in the osteoclast lineage (Extended Data Fig. 10a-c). *Tgfr2* expression was suppressed by miR-34a gain-of-function, but increased by miR-34a loss-of-function, both in mouse and human osteoclast cultures (Fig. 4a, b and Extended Data Fig. 10d, e). The miR-34a seed region in *Tgfr2* 3' untranslated region (UTR) is evolutionarily conserved in mammals (Fig. 4c). Luciferase reporter assay showed that *Tgfr2* 3' UTR is sufficient to confer miR-34a regulation (Fig. 4d, e). Importantly, when the miR-34a seed region in the *Tgfr2* 3' UTR was mutated, miR-34a regulation was abolished (Fig. 4d, e).

Tgif2 expression was increased during WT osteoclast differentiation (Fig. 4b). Tgif2-KO and Tgif2-Het mice had lower bone resorption and higher bone mass (Fig. 4f–h and Extended Data Fig. 10f). Tgif2 deletion reduced osteoclast differentiation, and abolished the anti-osteoclastogenic effects of miR-34a (Fig. 4i, j). Moreover, Tgif2/miR-34a double knockout mice (DKO) could no longer increase osteoclast differentiation or bone resorption (Fig. 4k, l) compared with Tgif2-KO mice. These results indicate that Tgif2 is pro-osteoclastogenic and essential for miR-34a regulation.

We next investigated how Tgif2 potentiates RANKL signalling. Transfection assays revealed that NFATc1, c-fos and c-jun, and to a lesser extent NF- κ B (p65), could induce Tgif2 expression (Fig. 4m). Response elements for NFATc1 and AP-1, but not NF- κ B, were identified in the Tgif2 promoter region (–5 to +5 kb). Chromatin immunoprecipitation analysis in osteoclast cultures showed that NFATc1, c-jun and c-fos bound to these sites upon RANKL stimulation, leading to activated Tgif2 transcription shown by the elevated H3K4me3 level at the transcription start site (Fig. 4n). This indicates that NFATc1 and AP-1 induce Tgif2 expression during osteoclastogenesis.

Luciferase reporter assay showed that Tgif2 augmented the activity of NFATc1, NF- κ B and c-Jun, but not c-fos (Fig. 4o). Consistently, the activity of endogenous NFATc1, NF- κ B and c-Jun, but not c-fos, was reduced in Tgif2-KO cultures and enhanced in 34a-KO cultures (Fig. 4p). Furthermore, NFATc1 messenger RNA (mRNA), c-Jun phosphorylation and I κ B α degradation were decreased in Tgif2-KO cultures and increased in 34a-KO cultures (Fig. 4q, r). Therefore, Tgif2 potentiates osteoclastogenesis via a positive feedback loop in which RANKL-induced transcription factors activate Tgif2 expression, and Tgif2 in turn promotes their activity. Collectively, these findings reveal Tgif2 as a novel yet critical regulator of osteoclastogenesis and bone resorption, as well as a key miR-34a direct target that is essential for miR-34a regulation (Fig. 4s).

The roles of miRNAs in bone physiology have just begun to emerge. Recent studies show that osteoblast-specific gain-of-function of miR-34b/c decreases bone mass by suppressing osteoblastogenesis and bone formation^{11,12}. Here we show that osteoclast-specific miR-34a gain-of-function protects bone by suppressing osteoclastogenesis and bone resorption. These findings uncover an interesting functional divergence among the miR-34 family members. Our study paves the road for future discovery of other miRNAs that may be regulated by RANKL and/or control Tgif2 expression, as well as future epidemiological and clinical studies to explore the pathological and therapeutic roles of this miR-34a-Tgif2 pathway in human.

Most systemically delivered drugs can target multiple tissues and cell types. We found that miR-34a also enhances bone formation; although miR-34a overexpression in osteoblast is neither sufficient nor essential for the therapeutic benefits of miR-34a in osteoporosis or bone metastasis, miR-34a may be a superior new therapy that exerts both anti-catabolic and anabolic effects compared with the current drugs that are solely anti-catabolic. Our identification of miR-34a, as well as the recent report of miR-141 and miR-219 (ref. 13), opens an exciting avenue for the development of a new generation of RNA-based osteo-protective medicine.

Other miR-34a target genes have been reported in different biological context, such as *SIRT1* (ref. 14), *SIRT6* (ref. 15) and *PNUTS*¹⁶. Although miR-34a may also target genes other than Tgif2 in osteoclasts, our genetic rescue *ex vivo* and *in vivo* shows that Tgif2 is the key miR-34a target, suggesting that other genes are probably secondary or functionally irrelevant to osteoclastogenesis.

miR-34a, miR-34b and miR-34c are commonly deleted in human cancers¹⁷. *In vitro* studies suggest that they may be critical mediators of p53 function and potential tumour suppressors¹⁸. Surprisingly, *in vivo* studies reveal that miR-34abc triple knockout mice exhibit intact p53 function without increased tumorigenesis¹⁰. Nonetheless, systemic miR-34a administration can indeed attenuate cancer malignancy¹⁹. This raises the intriguing possibility that its anti-cancer effects may reside in other cells that constitute the tumour microenvironment such as the osteoclasts in the bone metastatic niche. Indeed, our findings illustrate that bone

metastases are effectively blocked by miR-34a in osteoclasts, thus providing the first *in vivo* genetic evidence that miR-34a opposes malignant progression of cancer by disarming the metastatic niche.

METHODS SUMMARY

Conditional miR-34a transgenic mice were generated using the CAG-Z-EGFP vector. miR-34a knockout mice were generated using a gene-trap embryonic stem cell line.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 July 2013; accepted 8 April 2014.

Published online 25 June 2014.

1. Coleman, R. E. Bone cancer in 2011: prevention and treatment of bone metastases. *Nature Rev. Clin. Oncol.* **9**, 76–78 (2012).
2. Ell, B. & Kang, Y. SnapShot: bone metastasis. *Cell* **151**, 690 (2012).
3. Novack, D. V. & Teitelbaum, S. L. The osteoclast: friend or foe? *Annu. Rev. Pathol.* **3**, 457–484 (2008).
4. Chivukula, R. R. & Mendell, J. T. Circular reasoning: microRNAs and cell-cycle control. *Trends Biochem. Sci.* **33**, 474–481 (2008).
5. Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**, 586–591 (2009).
6. Kasinski, A. L. & Slack, F. J. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nature Rev. Cancer* **11**, 849–864 (2011).
7. Wan, Y., Chong, L. W. & Evans, R. M. PPAR- γ regulates osteoclastogenesis in mice. *Nature Med.* **13**, 1496–1503 (2007).
8. Wei, W. *et al.* PGC1 β mediates PPAR γ activation of osteoclastogenesis and rosiglitazone-induced bone loss. *Cell Metab.* **11**, 503–516 (2010).
9. Choi, Y. J. *et al.* miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nature Cell Biol.* **13**, 1353–1360 (2011).
10. Concepcion, C. P. *et al.* Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet.* **8**, e1002797 (2012).
11. Bae, Y. *et al.* miRNA-34c regulates Notch signaling during bone development. *Hum. Mol. Genet.* **21**, 2991–3000 (2012).
12. Wei, J. *et al.* miR-34s inhibit osteoblast proliferation and differentiation in the mouse by targeting SATB2. *J. Cell Biol.* **197**, 509–521 (2012).
13. Ell, B. *et al.* Tumor-induced osteoclast miRNA changes as regulators and biomarkers of osteolytic bone metastasis. *Cancer Cell* **24**, 542–556 (2013).
14. Yamakuchi, M., Ferlito, M. & Lowenstein, C. J. miR-34a repression of SIRT1 regulates apoptosis. *Proc. Natl Acad. Sci. USA* **105**, 13421–13426 (2008).
15. Lefort, K. *et al.* A miR-34a-SIRT6 axis in the squamous cell differentiation network. *EMBO J.* **32**, 2248–2263 (2013).
16. Boon, R. A. *et al.* MicroRNA-34a regulates cardiac ageing and function. *Nature* **495**, 107–110 (2013).
17. Lodygin, D. *et al.* Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* **7**, 2591–2600 (2008).
18. Hermeking, H. p53 enters the microRNA world. *Cancer Cell* **12**, 414–418 (2007).
19. Liu, C. *et al.* The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nature Med.* **17**, 211–215 (2011).

Acknowledgements We thank University of Texas Southwestern transgenic core and small animal imaging core for their assistance in our studies; P. Dechow, J. Feng and C. Qin for assistance with micro-computed tomography, histomorphometry and X-ray analysis; A. Ventura for miR-34abc triple knockout mice; D. Wotton for Tgif2-KO mice; H. Kronenberg for Osx-CreER mice; Y. Mishina for CAG-Z-EGFP vector. Y. Wan is a Virginia Murchison Linthicum Scholar in Medical Research. This work was in part supported by CPRIT (RP130145, Y.W.; R1008, J.M.), DOD (BC122877, Y.W.), National Institutes of Health (R01 DK089113, Y.W.; R01 CA120185 and P01 CA134292, J.M.; U54 CA151668 and UH2 TR000943, A.S.; R01 CA139067, L.H.), The Welch Foundation (I-1751, Y.W.) and a University of Texas Southwestern Endowed Scholar Startup Fund (Y.W.). The University of Texas Southwestern Small Animal Imaging Resource is supported in part by the Harold C. Simmons Cancer Center through an NCI Cancer Center Support Grant (1P30 CA142543) and The Department of Radiology. The VisualSonics Vevo 770 was purchased with National Institutes of Health American Recovery and Reinvestment Act stimulus funds 1S10R02564801.

Author Contributions J.Y.K. and Y.W. conceived the project and designed the experiments. All experiments, except those listed below, were performed by J.Y.K. W.W. assisted with micro-computed tomography, enzyme-linked immunosorbent assay and histomorphometry analyses. H.D.H. assisted with bone marrow transplantation and injection. Z.J. assisted with FACS analyses. X.W. assisted with western blot analyses. T.C.C. assisted with northern blot analyses and lifespan experiments. X.J.X. assisted with statistical analyses. L.H. provided the full miR-34a knockout mice. L.S.M., G.L.B. and A.K.S. assisted with nanoparticle packaging. J.T.M. provided the miR-34a gene trap knockout mice. Y.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.W. (yihong.wan@utsouthwestern.edu).

The long-term maintenance of a resistance polymorphism through diffuse interactions

Talia L. Karasov^{1,2*}, Joel M. Kniskern^{1,4*}, Liping Gao¹, Brody J. DeYoung³, Jing Ding^{1,†}, Ullrich Dubiella³, Ruben O. Lastra¹, Sumitha Nallu¹, Fabrice Roux^{4,5,6}, Roger W. Innes³, Luke G. Barrett^{1,†}, Richard R. Hudson¹ & Joy Bergelson¹

Plant resistance (*R*) genes are a crucial component in plant defence against pathogens¹. Although *R* genes often fail to provide durable resistance in an agricultural context, they frequently persist as long-lived balanced polymorphisms in nature^{2–4}. Standard theory explains the maintenance of such polymorphisms through a balance of the costs and benefits of resistance and virulence in a tightly coevolving host–pathogen pair^{5,6}. However, many plant–pathogen interactions lack such specificity⁷. Whether, and how, balanced polymorphisms are maintained in diffusely interacting species⁸ is unknown. Here we identify a naturally interacting *R* gene and effector pair in *Arabidopsis thaliana* and its facultative plant pathogen, *Pseudomonas syringae*. The protein encoded by the *R* gene *RPS5* recognizes an AvrPphB homologue (AvrPphB2) and exhibits a balanced polymorphism that has been maintained for over 2 million years (ref. 3). Consistent with the presence of an ancient balanced polymorphism, the *R* gene confers a benefit when plants are infected with *P. syringae* carrying *avrPphB2* but also incurs a large cost in the absence of infection. *RPS5* alleles are maintained at intermediate frequencies in populations globally, suggesting ubiquitous selection for resistance. However, the presence of *P. syringae* carrying *avrPphB* is probably insufficient to explain the *RPS5* polymorphism. First, *avrPphB* homologues occur at very low frequencies in *P. syringae* populations on *A. thaliana*. Second, AvrPphB only rarely confers a virulence benefit to *P. syringae* on *A. thaliana*. Instead, we find evidence that selection for *RPS5* involves multiple non-homologous effectors and multiple pathogen species. These results and an associated model suggest that the *R* gene polymorphism in *A. thaliana* may not be maintained through a tightly coupled interaction involving a single coevolved *R* gene and effector pair. More likely, the stable polymorphism is maintained through complex and diffuse community-wide interactions.

Individuals within a population often exhibit a wide range of susceptibilities to infection and this variability has a central influence on the emergence and spread of disease^{9–11}. The mechanisms that maintain resistance polymorphisms within populations have therefore been a major focus both of theoretical and empirical research^{12,13}. Despite this effort, significant gaps persist in our understanding of how and why genetic variation in resistance traits is maintained. For example, current theory on the long-term maintenance of resistance polymorphisms assumes that a pathogen specializes exclusively on one host species. However, roughly half of all plant pathogens associate with multiple hosts⁷, and many sparsely populated and ephemeral host species, such as *A. thaliana*, are rarely attacked by specialist pathogens. Despite the prevalence of generalist pathogens, ancient balanced polymorphisms in resistance are ubiquitous in *A. thaliana*². We currently lack an understanding of how these polymorphisms are maintained.

The interaction between *A. thaliana* and the model plant pathogen *P. syringae* is an ideal system to disentangle these dynamics both because *P. syringae* is a commonly occurring pathogen in wild populations of *A. thaliana*¹⁴ and because of the extensive knowledge of the molecular biology of this interaction¹. *P. syringae* interacts with *A. thaliana* through the secretion of virulence-associated proteins (effectors) that downregulate plant basal defence¹. *A. thaliana*, in turn, employs an arsenal of *R* gene products that recognize the action of specific effector proteins, inducing localized cell death (hypersensitive response) and a systemic production of chemical defences. These interactions fit the classic gene-for-gene model in requiring a close match between an effector and the corresponding *R* gene allele for a resistance response. Although *P. syringae* does not specialize on *A. thaliana*¹⁵, multiple *R* genes in *A. thaliana* recognize specific *P. syringae* effectors and exhibit balanced resistance polymorphisms that have been maintained for millions of years^{2–4}. Here we examine the dynamics of one specific gene-for-gene interaction in co-occurring *A. thaliana* and *P. syringae* populations to identify ecological and evolutionary processes underlying the long-term maintenance of resistance polymorphisms.

We first investigated the genetic basis of recognition in the interaction between *A. thaliana* and one of its resident *P. syringae* pathogens. *P. syringae* strain PNA29.1a was isolated from a natural *A. thaliana* population¹⁴ and shown to induce variable hypersensitive response on different *A. thaliana* accessions. To identify the effector eliciting hypersensitive response, we expressed cosmid clones of the PNA29.1a genome in a *P. syringae* strain, PstDC3000 (DC3000), which does not elicit hypersensitive response in *A. thaliana*. A genomic fragment conferring induction of hypersensitive response to DC3000 was found to contain a homologue of the *P. syringae* effector *avrPphB* (also known as *hopAR1* and *avrPph3*). Through subcloning, we confirmed that this effector, which we name AvrPphB2, is required for recognition. To identify the corresponding *R* gene, we mapped the hypersensitive response phenotype induced by DC3000(AvrPphB2) on an association panel of 75 *A. thaliana* accessions¹⁶ (Supplementary Table 1) and observed a single significant peak on chromosome 1 (Fig. 1a). The three most significant single nucleotide polymorphisms ($P = 1.27 \times 10^{-12}$) fell within 3.5 kilobases (kb) of *RPS5*, which encodes an R protein known to recognize AvrPphB¹⁷. To confirm that *RPS5* is responsible for recognition of AvrPphB2, we generated six isogenic pairs of resistant (*RPS5*⁺) and susceptible (*RPS5*[−]) plant lines and tested whether an AvrPphB2 induced hypersensitive response is restricted to the *RPS5*⁺ hosts. All *RPS5*⁺ lines expressed hypersensitive response in response to PNA29.1a after 16–24 h, whereas their *RPS5*[−] isolines did not. As expected, growth of DC3000(*avrPphB2*) was significantly reduced in *RPS5*⁺ plants relative to *RPS5*[−] plants, even after correcting for multiple tests ($P = 1.45 \times 10^{-3}$, Wilcoxon rank-sum test) (Extended Data Fig. 1).

¹Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. ²Committee on Genetics Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA.

³Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. ⁴INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, F-31326 Castanet-Tolosan, France. ⁵CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, F-31326 Castanet-Tolosan, France. ⁶Laboratoire Génétique et Evolution des Populations Végétales, UMR CNRS 8198, Université des Sciences et Technologies de Lille – Lille 1, F-59655 Villeneuve d'Ascq Cedex, France. [†]Present addresses: Monsanto Vegetable Seeds, 37437 State Highway 16, Woodland, California 95695, USA (J.M.K.); State Key Laboratory of Plant Genetics and Germplasm Enhancement and College of Horticulture, Nanjing Agricultural University, Nanjing, China (J.D.); CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia (L.G.B.).

*These authors contributed equally to this work.

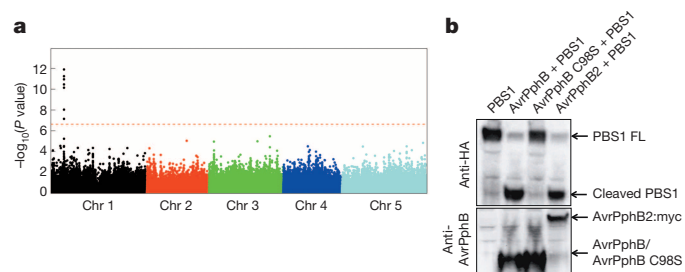


Figure 1 | Identification of RPS5 and AvrPphB2 as a naturally interacting R-gene-effector pair in *A. thaliana* and *P. syringae* populations.

a, Hypersensitive response in response to infection by DC3000 (*avrPphB2*) was scored in 75 worldwide *A. thaliana* accessions that were previously genotyped at ~250,000 single nucleotide polymorphisms¹⁶. EMMAX³⁰ was used to perform genome-wide association mapping of differential hypersensitive response. The Manhattan plot illustrates the *P* values associated with each of the single nucleotide polymorphisms, and the dotted line signifies the Bonferroni correction threshold for significance ($P = 2.33 \times 10^{-7}$). The top three single nucleotide polymorphisms, each with *P* values 1.27×10^{-12} , lie within 3.5 kb of *RPS5* on chromosome (Chr) 1. **b**, AvrPphB2 cleaves PBS1. Immunoblot of PBS1 tagged with haemagglutinin (PBS1-HA) expressed alone or co-expressed with AvrPphB, AvrPphB C98S (non-active mutant) or AvrPphB2-myc in *Nicotiana benthamiana*. AvrPphB and AvrPphB2 both cleave the full-length (FL) PBS1 whereas AvrPphB C98S does not. The upper panel shows the immunoblotting results for the HA-tagged PBS1 and the bottom panel shows the results for immunoblotting for AvrPphB variants.

AvrPphB2 exhibits relatively low identity with AvrPphB at both nucleotide and protein levels (75% and 78% respectively). This divergence prompted us to ask whether AvrPphB2 triggers host resistance in a manner similar to AvrPphB by testing whether AvrPphB2 also cleaves the host protein PBS1, a virulence target of AvrPphB¹⁸. An immunoblot assay demonstrated cleavage of PBS1 by AvrPphB2 (Fig. 1b). Furthermore, infection of *RPS5*⁺/*PBS1*⁻ plants with DC3000 (*avrPphB2*) failed to elicit the hypersensitive response, confirming a conserved mode of action for PBS1-dependent RPS5 recognition of AvrPphB2. Thus, we find no evidence that sequence divergence in AvrPphB2 is associated with functional divergence in cleavage of PBS1.

RPS5 was previously shown to exhibit an ancient balanced polymorphism for its presence or absence in *A. thaliana*³. We confirmed this finding through a comparison of sequence divergence at the intergenic region surrounding the *RPS5* locus and here estimate that the polymorphism has been maintained in *A. thaliana* for approximately 2.6 million years (with 95% confidence intervals of 1.5–3.9 million years (refs 19, 20)). As shown previously³, the divergences between the *Arabidopsis lyrata* *RPS5* locus and the *A. thaliana* presence and absence alleles are equivalent, indicating that the deletion event did not skew the inference of polymorphism age. Furthermore, we extended this work to examine polymorphisms within a broad array of *A. thaliana* populations worldwide using a PCR assay for *RPS5* presence/absence. The survey of 1,198 genetically distinct plants from 357 populations revealed that *RPS5* was present in 51% of individuals (Fig. 2). Within the 39 populations in which we detected four or more distinct genotypes, *RPS5* was found at an average frequency of 56% (Fig. 2a), with both alleles present in more than 90% of these populations. Thus, *RPS5* polymorphism is long-lived, common in extant populations and maintained at intermediate frequency in many local populations.

This ongoing maintenance of an ancient R gene polymorphism within populations suggests strongly that *RPS5* confers both a fitness advantage and a fitness cost²¹, probably dependent upon the presence or absence of specific pathogen effector loci. To test for a cost of resistance in the absence of infection, we measured the relative fitness of *RPS5*⁺/*RPS5*⁻ isolines in a field experiment performed in the absence of *avrPphB* homologues, including *avrPphB2*. Our experimental design included six pairs of *RPS5*⁺/*RPS5*⁻ isolines, with replicates of each pair randomized within blocks in a field in Downer's Grove, Illinois, USA. Importantly, we used *RPS5*⁺/*RPS5*⁻

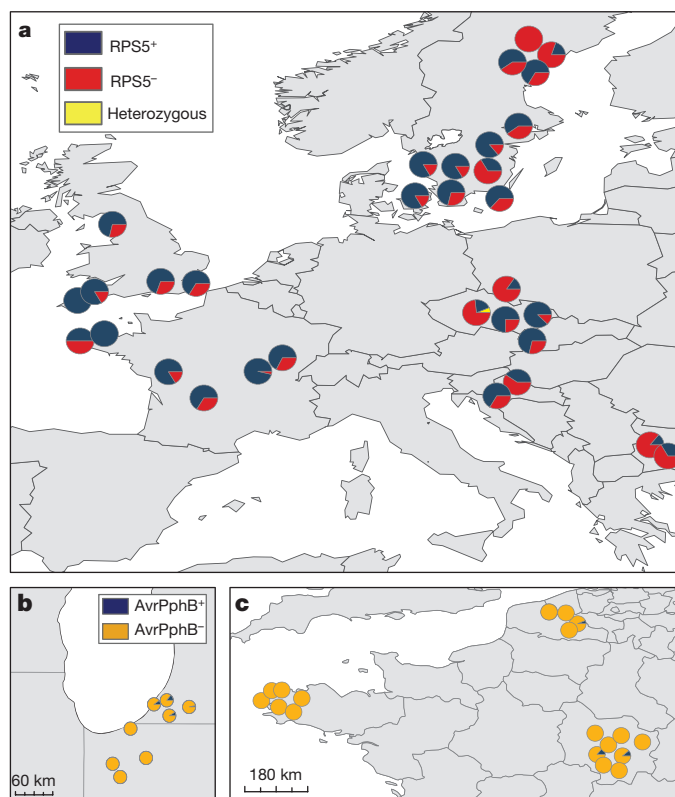


Figure 2 | The distribution of *RPS5* and *avrPphB* homologues in co-occurring *A. thaliana* and *P. syringae* populations. PCR was used to test for the frequency of *RPS5* in *A. thaliana* populations across Europe (**a**) and the Midwestern USA (Extended Data Fig. 3). The *RPS5* locus was polymorphic in more than 90% (36 out of 39) of these populations (one population in eastern Asia is not illustrated). Dot-blot assays tested for the presence of AvrPphB homologues in isolates of *P. syringae* from the Midwestern USA (**b**) and from France (**c**). *avrPphB* homologues are found at 3.27% frequency in the Midwestern USA (*RPS5* frequency ~24%) and at 2.29% frequency in France (*RPS5* frequency ~81%).

isolines generated both in susceptible and resistant backgrounds (via Cre-lox recombination and EMS mutagenesis respectively) to allow assessment of fitness effects due to autoimmunity²². For the isogenic *RPS5*^{+/−} pairs in the susceptible and resistant backgrounds respectively, the *RPS5*⁺ genotype produced fewer seeds ($P = 0.002$, $P < 0.001$), fewer siliques ($P = 0.002$, $P < 0.001$) and less plant biomass ($P < 0.001$ both) than the *RPS5*⁻ genotype (paired *t*-test, $n = 214$ in the susceptible background and $n = 478$ in the resistant background). The fitness cost associated with the *RPS5*⁺ allele ranged from 5 to 10.2% (Fig. 3).

The large fitness cost associated with the presence of *RPS5* suggests that a high intensity of pathogen-mediated selection is required for the resistance polymorphism to be maintained. Consistent with this expectation, a fitness benefit of roughly 20% has been reported for plants carrying *RPS5* when infected with *P. syringae* containing *avrPphB2*²³. In contrast, a survey examining the prevalence of *avrPphB* and its homologues in *P. syringae* isolates from *A. thaliana* plants in the Midwestern USA revealed that *avrPphB* homologues were restricted to 4 out of 11 populations tested, and maintained at very low frequency (6 of 183 isolates (3.27%); Fig. 2b), despite *RPS5* being found at intermediate frequencies in this region (24% of plants in US populations; Extended Data Fig. 2). *A. thaliana* was introduced relatively recently into North America, and the ancient balanced polymorphism predates this introduction. To determine whether frequencies of *avrPphB* homologues are greater in Europe where it is naturalized, we tested for the presence of AvrPphB homologues in 131 isolates from the extended *P. syringae* species complex¹⁵ from 19 *A. thaliana* populations across France (Fig. 2c). We found putative distant homologues in three (2.29%) of these isolates.

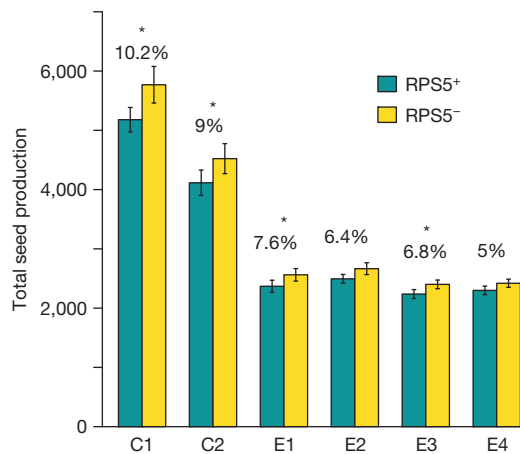


Figure 3 | The cost of RPS5-mediated resistance in the absence of AvrPphB homologues. Field trials of paired isogenic lines of *A. thaliana* that differed only in the presence or absence of RPS5 driven by its natural promoter. Lines C1 and C2 were generated by the insertion of RPS5 and its flanking regions into the susceptible Ga-0 background and subsequent excision of RPS5 using a Cre-lox system. Lines E1–E4 were generated by random point mutation in the native RPS5 gene to generate a null allele in the resistant Col-0 background (described in Supplementary Information). In all pairings, RPS5⁺ plants exhibited reduced fitness relative to RPS5[−] plants, ranging in magnitude from 5.0 to 10.2%. The percentage indicates the percentage decrease in seed production in the RPS5⁺ line relative to the RPS5[−] line, and the stars denote $P < 0.05$ in a paired t -test (with 51 and 56 plant pairings in the two susceptible backgrounds, and 59, 52, 66 and 62 plant pairings in the resistant backgrounds). Results are presented as the mean \pm one s.e.m.

Despite the rarity of AvrPphB homologues, RPS5⁺ is common in French populations (81% of plants in French *A. thaliana* populations possess RPS5). These results again suggest that a pairwise interaction between *A. thaliana* and *P. syringae* carrying homologues of AvrPphB may be insufficient for the maintenance of RPS5.

Current models to explain the maintenance of polymorphism in antagonistic species interactions typically assume tightly coupled interactions between specific host and enemy genotypes^{5,6}. However, single *R* genes may recognize and interact with multiple pathogen effectors^{7,24}. We hypothesized that the high frequency of RPS5 within *A. thaliana* populations may result from selection imposed by effectors other than AvrPphB homologues. We tested this idea by examining differential hypersensitive responses between RPS5^{+/−} isogenic plants when infected with 42 *A. thaliana*-isolated *P. syringae* strains from the Midwestern USA. These strains were chosen to lack AvrPphB and its homologues on the basis of dot-blot hybridizations. One strain differentially induced hypersensitive response on RPS5⁺ plants, suggesting that RPS5 may encode a protein that has the capacity to recognize non-homologous effectors (Supplementary Table 2). However, the low frequency of *P. syringae* isolates with any effector recognized by RPS5 in these 42 isolates (<5%) suggests additional sources of selection for resistance.

A. thaliana is host to a diverse bacterial community within its leaves²⁵. A selective advantage for RPS5 could result if multiple *A. thaliana* pathogens carry AvrPphB homologues or non-homologous effectors recognized by RPS5. In support of the former, sequence analysis suggests the potential for horizontal transfer of AvrPphB among bacterial species. A comparison of the genomes of a *P. syringae* kiwi crop pathovar and an *A. thaliana* pathovar revealed that synonymous divergence between AvrPphB and AvrPphB2 is substantially greater than divergence between other shared homologues (synonymous divergence = 2.28 versus an average synonymous divergence of 0.60; empirical $P = 0.003$) (Extended Data Fig. 3). The unusually ancient coalescence between AvrPphB homologues raises the possibility that at least one AvrPphB homologue was horizontally transferred from another bacterial species. If AvrPphB homologues persist in another species, interactions between *A. thaliana* and this other species could contribute to selection for RPS5.

Spillover of pathogens from alternative neighbouring hosts has been shown to influence resistance traits in plant populations⁷. In the agricultural and disturbed ecosystems where *A. thaliana* is commonly found, *Arabidopsis* exists as a small ephemeral that is a less predictable and productive host than co-occurring perennials and crops. Other hosts can provide strong selection on the virulence of *P. syringae*; indeed, isolates of *P. syringae* collected from other hosts consistently grow in *A. thaliana* where they can demonstrate unusually high virulence²⁶. Infrequent infection by highly virulent strains of *P. syringae* could impose strong selection for the ability to recognize these pathovars. It is unknown, however, if RPS5 recognizes effectors in a wide breadth of pathovars.

To determine whether AvrPphB and its homologues are present in strains isolated from alternative hosts, we probed a panel of 78 *P. syringae* strains isolated from crops. Successful hybridization to an AvrPphB probe was observed for 21 out of 78 (27%) of the pathovar strains (Supplementary Table 3). We cloned 14 of these homologues, expressed them in DC3000 and tested for RPS5-mediated recognition. These homologues were chosen to span diverse *P. syringae* clades and thus provide a representative sample of how *A. thaliana* and *P. syringae* interact at the RPS5/avrPphB interface, at least in an agricultural context. Of course, selection imposed by crop pathovars cannot explain a polymorphism that predates the advent of agriculture. Nevertheless, these pathovars provide a window into how effectors carried by non-resident strains interact with *A. thaliana*. Approximately 70% (10 out of 14) of these homologues were recognized, indicating that non-resident *P. syringae* can be subject to RPS5-mediated resistance (Extended Data Fig. 4). Although the frequency of pathogenic spillover onto *A. thaliana* populations is not known, this result demonstrates that pathovars of *P. syringae* have the potential to impose selection favouring the presence of RPS5 in *A. thaliana* populations.

The low frequency of AvrPphB in *P. syringae* residing within the leaves of *A. thaliana* suggests a limited fitness advantage for carrying this gene. We compared growth of DC3000 carrying an empty plasmid with the growth of DC3000(avrPphB2) (Extended Data Fig. 5), and found that AvrPphB2 significantly enhanced growth within an RPS5[−] accession (Wilcoxon rank-sum test, $P = 0.04$). The common laboratory strain DC3000 is a crop pathovar highly diverged from most *A. thaliana* isolates of *P. syringae*²⁷; consequently, the virulence effect of AvrPphB2 may differ in the endemic *P. syringae* isolates. Similar experiments using each of three *P. syringae* isolates collected from *A. thaliana* populations revealed a virulence benefit in only one of the three isolates (Extended Data Fig. 6) (Wilcoxon rank-sum tests; $P = 0.01, 0.90$ and 0.50 respectively). Virulence benefits that are dependent on the genetic background help explain the relatively low frequency of AvrPphB2 within *A. thaliana* populations.

These observations are not consistent with the tight *R* gene and effector pairing thought necessary to generate stable resistance polymorphisms^{4,12}. Our data instead suggest that selection favouring RPS5 is diffuse, deriving from multiple and perhaps individually small selective agents. Similarly, the frequency of the effector in the pathogen may be only weakly coupled to the frequency of RPS5 in *A. thaliana*. This is particularly true if *P. syringae* isolates intermittently and transiently spill over from alternative hosts to infect *A. thaliana* populations. This inconsistency between longstanding theory and our observations raises the possibility that diffuse interactions between *A. thaliana* and *P. syringae* may generate stable balanced polymorphisms.

To test this possibility, we modelled the dynamics of a resistance polymorphism under a range of parameters for a diffusely interacting host–pathogen pair. More specifically, we asked whether a balanced polymorphism could be maintained when the frequency of an *R* gene polymorphism depends upon the frequency of recognized effectors, but the frequency of effectors is uncoupled from *R* gene dynamics. To determine the conditions under which the polymorphism is stably maintained (stability is defined here as maintenance $\geq 100,000$ generations), we modelled the dynamics of an *R* gene polymorphism across a range of values for the cost and the frequency of infection (Fig. 4). Next, we incorporated frequency dependence as a mechanism to couple effector and *R* gene dynamics asymmetrically without explicitly modelling effector

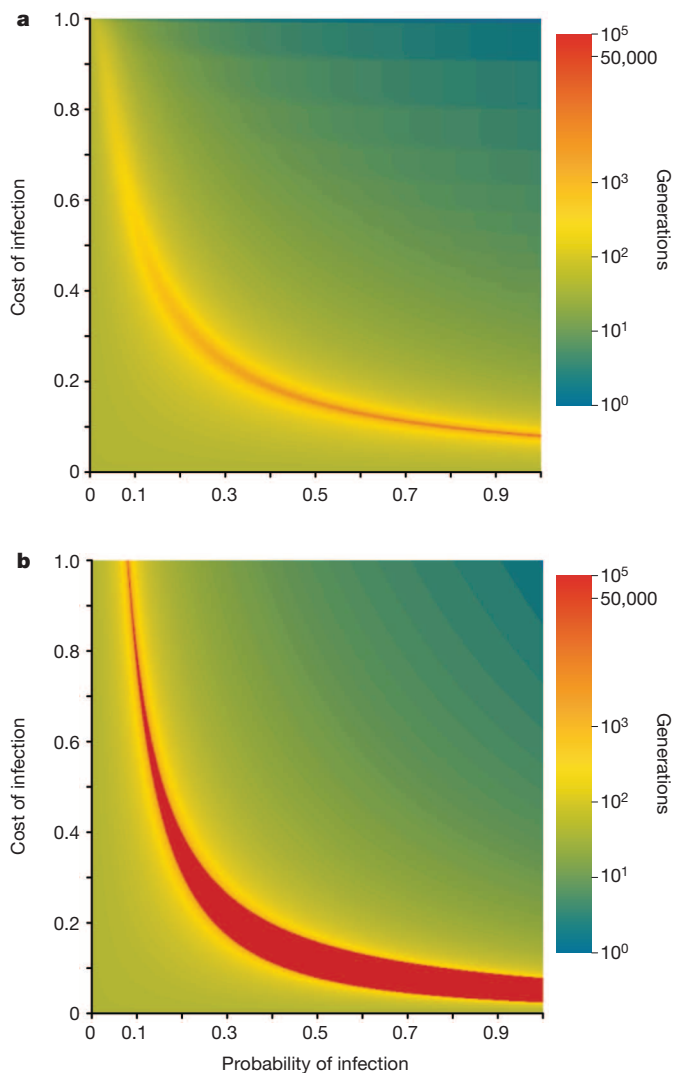


Figure 4 | The maintenance of a balanced polymorphism in a diffuse interaction. To ascertain the conditions sufficient to maintain an *R* gene polymorphism stably in a diffuse interaction, we modelled the dynamics of an *R* gene polymorphism with a range of parameters for the cost of infection (y axis) and probability of infection with bacteria containing a recognized effector (x axis). With a starting frequency of 0.5 for the *R* gene and a cost of resistance of 0.08 (a value consistent with our observations in Fig. 3), we recursively simulated the frequency of the *R* gene to determine the number of generations that the *R* gene remained at a putatively detectable frequency (between 0.01 and 0.99). The simulation was run for 100,000 generations. The colour bar signifies the number of generations, up to 100,000, that the polymorphism is maintained. For **a**, the dynamics were modelled without frequency dependence. The *R* gene polymorphism is stably maintained for a narrow range of parameters. With the incorporation of negative frequency dependence into the model (described in Supplementary Information), as illustrated in **b**, the parameter space allowing a stable polymorphism expands.

evolution (reviewed in ref. 12). We envision a situation in which infection of susceptible plants increases resource availability for uninfected plants owing to a population-wide reduction in plant biomass or numbers. The addition of frequency dependence substantially expands the parameter space over which the *R* gene polymorphism is maintained (Fig. 4b). Across a range of infection rates (20–99%), and for relatively low costs of infection (1–30%), the balanced polymorphism can be maintained for thousands of generations (Fig. 4b and Extended Data Fig. 7). Thus, the maintenance of resistance polymorphisms does not require a tight coevolutionary pairing.

Even though it is unlikely that many *A. thaliana*–pathogen interactions are tightly coupled (but see ref. 28), *A. thaliana* populations exhibit

widespread balancing selection at *R* gene loci². Little is known of the ecological factors that drive *R* gene evolution in *A. thaliana*, and recent studies suggest that selection on resistance is both pathogen and host dependent. Our results point to the maintenance of one of these ancient balanced polymorphisms through complex interactions with multiple hosts sharing *P. syringae*, and through *RPS5* interacting with multiple effectors distributed across multiple pathogen species. Recent studies revealing *R*-gene-mediated sexual incompatibilities and divergent selection for resistance to necrotrophs versus biotrophs suggest that numerous additional interactions may select for variation at resistance loci^{22,29}. Our corresponding model demonstrates that a stable polymorphism does not require the tight coevolutionary coupling of a host and pathogen, but rather can be maintained in the face of complex interactions such as the ones we observe. Many, if not most, pathogens persist in complex ecological communities, with multiple hosts and competitors⁷. Future work on host–pathogen coevolution should elucidate whether this complexity is a general mechanism for the maintenance of variation in susceptibility and resistance.

METHODS SUMMARY

A genomic library of PNA29.1a was created using the broad-host-range vector pLAFR1, and the genomic fragment containing *avrPphB2* expressed in PstDC3000 was shown to be sufficient to induce hypersensitive response in Col-0 plants. The genetic basis of differential hypersensitive response in 75 *A. thaliana* ecotypes in response to infection with PstDC3000(*avrPphB2*) was mapped¹⁶ using EMMAX³⁰. The frequency of *RPS5* was determined by a PCR assay on genetically distinct *A. thaliana* lines. The fitness effects of the *RPS5* locus for *A. thaliana* were tested through comparisons of six paired isogenic lines. Two of the pairs were derived from Cre-lox site-specific recombination involving a fragment of the Col-0 ecotype containing the *RPS5* coding region, and the promoter and terminator regions in the Ga-0 ecotype. The remaining four pairs of isogenic lines within the Col-0 background were backcrosses of an *RPS5* mutant containing an early stop codon at amino-acid position 319 generated in the *Arabidopsis* TILLING project. To measure the fitness effects of *AvrPphB2*, the growth of a *P. syringae* strain transformed with an empty pME6010 plasmid was compared with strains transformed with pME6010 containing *avrPphB2*. *avrPphB* homologues were amplified and cloned into a modified version of the Gateway-compatible vector pMTN41, then expressed in DC3000. *RPS5*^{+/–} plants were infected with DC3000 carrying the different homologues to detect differential hypersensitive response. The dynamics of the *R* gene polymorphism over time were modelled with an infinite population size and a cost of resistance of 8%. For frequency-dependent selection, the fitness of a resistant individual was modelled as increasing with the frequency of susceptible individuals.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 November 2012; accepted 1 May 2014.

Published online 6 July 2014.

- Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
- Bakker, E. G., Toomajian, C., Kreitman, M. & Bergelson, J. A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell* **18**, 1803–1818 (2006).
- Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11525–11530 (2002).
- Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671 (1999).
- Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85**, 411–426 (1982).
- Frank, S. A. Models of plant–pathogen coevolution. *Trends Genet.* **8**, 213–219 (1992).
- Barrett, L. G., Kniskern, J. M., Bodenhausen, N., Zhang, W. & Bergelson, J. Continuum of specificity and virulence in plant host–pathogen interactions: causes and consequences. *New Phytol.* **183**, 513–529 (2009).
- Stinchcombe, J. R. & Rausher, M. D. Diffuse selection on resistance to deer herbivory in the ivyleaf morning glory, *Ipomoea hederacea*. *Am. Nat.* **158**, 376–388 (2001).
- Lockett, S. F. et al. Mismatched human leukocyte antigen alleles protect against heterosexual HIV transmission. *J. Acquir. Immune Defic. Syndr.* **27**, 277–280 (2001).
- Paterson, S., Wilson, K. & Pemberton, J. M. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. *Proc. Natl Acad. Sci. USA* **95**, 3714–3719 (1998).

11. Thrall, P. H. *et al.* Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecol. Lett.* **15**, 425–435 (2012).
12. Brown, J. K. & Tellier, A. Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annu. Rev. Phytopathol.* **49**, 345–367 (2011).
13. Haldane, J. B. S. Disease and evolution. *Curr. Sci.* **63**, 599–604 (1992).
14. Jakob, K. *et al.* *Pseudomonas viridiflava* and *P. syringae*—natural pathogens of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* **15**, 1195–1203 (2002).
15. Morris, C. E. *et al.* Inferring the evolutionary history of the plant pathogen *Pseudomonas syringae* from its biogeography in headwaters of rivers in North America, Europe, and New Zealand. *MBio* **1**, e00107–10 (2010).
16. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
17. Warren, R. F., Merritt, P. M., Holub, E. & Innes, R. W. Identification of three putative signal transduction genes involved in *R* gene-specified disease resistance in *Arabidopsis*. *Genetics* **152**, 401–412 (1999).
18. Shao, F. *et al.* Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301**, 1230–1233 (2003).
19. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97**, 7360–7365 (2000).
20. Hudson, R. R. The variance of coalescent time estimates from DNA sequences. *J. Mol. Evol.* **64**, 702–705 (2007).
21. Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
22. Bomblies, K. *et al.* Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5**, e236 (2007).
23. Gao, L., Roux, F. & Bergelson, J. Quantitative fitness effects of infection in a gene-for-gene system. *New Phytol.* **184**, 485–494 (2009).
24. Grant, M. R. *et al.* Structure of the *Arabidopsis* Rpm1 gene enabling dual-specificity disease resistance. *Science* **269**, 843–846 (1995).
25. Bodenhausen, N., Horton, M. W. & Bergelson, J. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* **8**, e56329 (2013).
26. Kniskern, J. M., Barrett, L. G. & Bergelson, J. Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution* **65**, 818–830 (2011).
27. Barrett, L. G., Bell, T., Dwyer, G. & Bergelson, J. Cheating, trade-offs and the evolution of aggressiveness in a natural pathogen population. *Ecol. Lett.* **14**, 1149–1157 (2011).
28. Allen, R. L. *et al.* Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957–1960 (2004).
29. Lorang, J. *et al.* Tricking the guard: exploiting plant defense for disease susceptibility. *Science* **338**, 659–662 (2012).
30. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348–354 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Dwyer, M. Kreitman, L. Merwin, C. Morris and M. Nordborg for discussions, D. Baltrus, D. Dahlbeck, J. Greenberg and B. Vinatzer for supplying plasmids, N. Faure for help collecting the French isolates of *P. syringae*, C. Godé, J. Higgins and A. Stathos for contributions to genotyping *RPS5* presence/absence, G. Sperone for mapping, and T. Morton and the staff of the University of Chicago greenhouse for planting and maintaining *A. thaliana*. T.L.K. was supported by a Department of Education GAANN fellowship, J.M.K. and L.G.B. were supported by postdoctoral fellowships from the Dropkin Foundation, and R.L. was supported by a National Institutes of Health (NIH) Postbaccalaureate Research Education Program award. F.R. was supported by the Laboratory of Excellence (Labex) TULIP (ANR-10-LABX-41; ANR-11-IDEX-0002-02). This work was supported by grants NIH-NIGMS R01 GM046451 to R.W.I., and grants NSF MCB0603515, NIH-NIGMS R01 GM057994 and NIH-NIGMS R01 GM083068 to J.B.

Author Contributions J.B. conceived the project and organized components; J.M.K. and R.L. cloned AvrPphB2 and performed functional analyses, J.M.K. mapped and cloned *RPS5*; L.G. performed fitness trials; B.J.D. and U.D. performed effector biochemical analyses; T.L.K. performed sequence divergence analyses; J.M.K., F.R. and J.D. assayed *RPS5* frequencies in *A. thaliana*; J.M.K., S.N. and T.L.K. performed *P. syringae* fitness assays; T.L.K. cloned AvrPphB homologues; T.L.K. and R.R.H. designed the population genetics analyses and R.W.I. designed the biochemical analyses; T.L.K. performed the modelling; J.B., L.B., T.L.K. and J.M.K. were involved in the study design, the experiments and the analyses; T.L.K. and J.B. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. (jbergs@uchicago.edu).

The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase

Fei Wang¹, Charlene Chan¹, Nicholas R. Weir¹ & Vladimir Denic¹

Hundreds of tail-anchored proteins, including soluble *N*-ethylmaleimide-sensitive factor attachment receptors (SNAREs) involved in vesicle fusion, are inserted post-translationally into the endoplasmic reticulum membrane by a dedicated protein-targeting pathway^{1–4}. Before insertion, the carboxy-terminal transmembrane domains of tail-anchored proteins are shielded in the cytosol by the conserved targeting factor Get3 (in yeast; TRC40 in mammals)^{5–7}. The Get3 endoplasmic-reticulum receptor comprises the cytosolic domains of the Get1/2 (WRB/CAML) transmembrane complex, which interact individually with the targeting factor to drive a conformational change that enables substrate release and, as a consequence, insertion^{8–11}. Because tail-anchored protein insertion is not associated with significant translocation of hydrophilic protein sequences across the membrane, it remains possible that Get1/2 cytosolic domains are sufficient to place Get3 in proximity with the endoplasmic-reticulum lipid bilayer and permit spontaneous insertion to occur^{12,13}. Here we use cell reporters and biochemical reconstitution to define mutations in the Get1/2 transmembrane domain that disrupt tail-anchored protein insertion without interfering with Get1/2 cytosolic domain function. These mutations reveal a novel Get1/2 insertase function, in the absence of which substrates stay bound to Get3 despite their proximity to the lipid bilayer; as a consequence, the notion of spontaneous transmembrane domain insertion is a non sequitur. Instead, the Get1/2 transmembrane domain helps to release substrates from Get3 by capturing their transmembrane domains, and these transmembrane interactions define a bona fide pre-integrated intermediate along a facilitated route for tail-anchor entry into the lipid bilayer. Our work sheds light on the fundamental point of convergence between co-translational and post-translational endoplasmic-reticulum membrane protein targeting and insertion: a mechanism for reducing the ability of a targeting factor to shield its substrates enables substrate handover to a transmembrane-domain-docking site embedded in the endoplasmic-reticulum membrane.

We have previously found that elution of substrates from Get3 immobilized on a resin can be achieved in the absence of any membranes by the addition of an engineered heterodimer of Get1/2 cytosolic domains (miniGet1/2)⁸. At physiological protein concentrations, mini-Get1/2 enabled substrate elution in a manner that was dependent on the interactions of both Get1/2 cytosolic domains with Get3. Notably, substrate elution by miniGet1/2 was also dependent on the presence of an engineered tail-anchored trap derived from Sgt2, a transmembrane domain (TMD)-recognition factor that delivers newly synthesized tail-anchored proteins to Get3 (ref. 14). By chemical crosslinking between Get3 and substrate, we have subsequently found that the tail-anchored trap prevents apparent re-binding of substrates to Get3 (Extended Data Fig. 1a). Thus, in the simplest model for insertion, the only role of the Get1/2 TMD is to physically link the Get1/2 cytosolic domains so that they can work together to enable ‘trapping’ of substrate tail anchors by the nearby hydrophobic lipid bilayer. A more complex alternative to this spontaneous insertion model is that the Get1/2 TMD is an insertase that facilitates entry of substrate tail anchors into the lipid bilayer.

The spontaneous insertion model predicts that the insertion of Get3 substrates should be insensitive to genetic perturbations of the Get1/2 TMD, which mediates complex formation, as long as the function of the Get1/2 cytosolic domains is preserved. To avoid the potential for complex disruption by mutations in the six transmembrane segments (Get1 TM1–3 and Get2 TM1–3), we first engineered a single-chain version of the Get1/2 heterodimer (Get2-1sc) expressed from the endogenous *GET2* promoter in *Δget1* cells. The resulting protein fusion was functional (Extended Data Fig. 1b and Fig. 1a), as measured using a green fluorescent protein (GFP) cell reporter of heat-shock factor transcriptional activity¹⁵, which is a good monitor of tail-anchored protein aggregation in the cytosol due to compromised Get1/2 function⁶. Get1/2 transmembrane segments were replaced with transmembrane segments from unrelated endoplasmic reticulum (ER) membrane proteins, either Sec61-β or Ost4 (Extended Data Fig. 2a). In addition, we mutated an absolutely conserved aspartic acid residue near the middle of Get2 TM3 (D271K) because replacement of this transmembrane segment severely destabilized Get2-1sc (Extended Data Fig. 2b and data not shown). All the mutations in the Get1/2 TMD resulted in the loss of Get2-1sc function, as evidenced by elevated heat-shock factor activity, with some alleles resulting in more apparent heat shock than others (Fig. 1a).

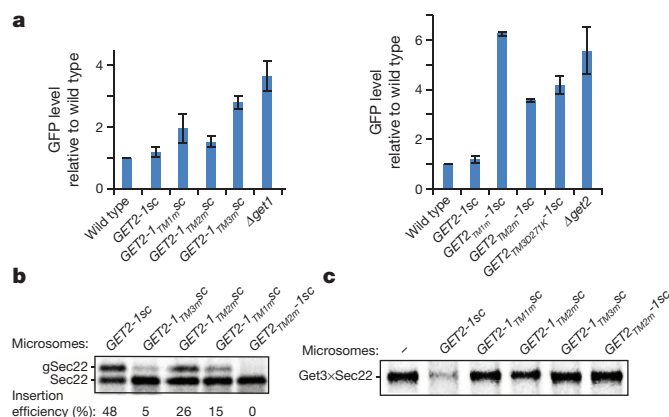


Figure 1 | In vivo and in vitro analysis of loss-of-function mutations in the Get1/2 TMD. **a**, The amount of GFP produced by the heat-shock reporter (Extended Data Fig. 1b) in the indicated strains was determined by flow cytometry analysis and plotted as the average of two independent experiments ($n = 2$). **b**, Affinity-purified Get3–Sec22 was incubated with the indicated microsomes at room temperature. Samples were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and visualized by autoradiography. The positions of unglycosylated and glycosylated Sec22 (gSec22) are indicated. Insertion efficiency is expressed as the percentage of Sec22 that is glycosylated. **c**, Affinity-purified Get3–Sec22 was incubated with the indicated microsomes or mock-incubated at room temperature followed by crosslinking with disuccinimidyl suberate (amine-reactive crosslinker). Samples were subjected to SDS–PAGE and visualized by autoradiography. The crosslinked product between Get3 and Sec22 is indicated (Get3 × Sec22).

¹Department of Molecular and Cellular Biology, Harvard University, Northwest Labs, Cambridge, Massachusetts 02138, USA.

To measure the impact of TMD mutations on Get1/2 activity more directly, we first produced radiolabelled Sec22 (a SNARE tail-anchored protein that facilitates vesicle fusion in the early secretory system) by *in vitro* translation in a wild-type budding yeast cell extract. Next, we affinity-purified Get3–Sec22 and monitored insertion into ER-derived membranes (microsomes) by glycosylation at a carboxy (C)-terminal glycan attachment site. We observed that loss-of-function mutations in the Get1/2 TMD also caused a proportional decrease in Sec22 insertion (Fig. 1b and Extended Data Fig. 3a). Chemical crosslinking between Get3 and Sec22 revealed that reduced Sec22 insertion was due to reduced substrate release from Get3 (Fig. 1c and Extended Data Fig. 3b). Even though Sec22 has a relatively hydrophobic TMD that might hinder its release from Get3, another tail-anchored protein, Sec61- β , which has a significantly less hydrophobic TMD, was similarly compromised by genetic disruptions of the Get1/2 TMD (Extended Data Fig. 3c).

Loss of Get1/2 TMD function would be expected to cause a defect in Get3 substrate release if the ER bilayer were by itself not an efficient ‘trap’ for TMDs delivered into its proximity by Get3 interactions with Get1/2 cytosolic domains. We performed two stringent tests of this idea by focusing on our strongest loss of function allele, Get2^{TM2m}-1sc, which did not apparently disrupt Get2-1sc ER protein targeting and membrane topology (Supplementary Discussion and Extended Data Figs 4 and 5). First, we took advantage of our previous observation that Get3 interacts more avidly with miniGet1/2 than with either miniGet1^{CDm}/2 or miniGet1/2^{CDm}, which contain point mutations in Get1 and Get2 cytosolic domains that abolish interactions with Get3 (ref. 8). This explains why Get3–Sec22 association with microsomes during a flotation assay depends on its interactions with both Get1/2 cytosolic domains⁸: avidity is necessary for targeting complexes to remain membrane-bound in the absence of Get3 re-binding during the 3 h in which microsomes float to the top of a density gradient. Consistent with our starting hypothesis, we found that Get3–Sec22 targeting complexes maintained a better association with Get2^{TM2m}-Get1sc microsomes relative to control Get2^{CDm}/Get1 microsomes (Fig. 2a). As a second test, we adapted our tail-anchored trap system to work on the surface of membranes by preparing proteoliposomes with three ingredients: a phospholipid mixture that approximated the composition of the ER membrane, trace

amounts of a Ni-NTA (nitrilotriacetic acid) phospholipid for subsequent tethering of the His₆-tagged tail-anchored trap, and Get2^{TM2m}-Get1sc that we affinity purified from yeast in complex with endogenous Get3 (Fig. 2b and Extended Data Fig. 6a). In further support of our hypothesis, we observed that the tail-anchored trap increased substrate release from Get2^{TM2m}-Get1sc proteoliposomes (Fig. 2b). Importantly, the tail-anchored trap was not able to rescue the substrate release defect of Get2/Get1^{CDm} proteoliposomes (Fig. 2b and Extended Data Fig. 6b). In summary, our data demonstrate that the ER bilayer is by itself not an efficient ‘trap’ for tail-anchored proteins delivered to its proximity by Get3 interactions with the cytosolic domains of Get1/2. They also argue that the Get1/2 complex has a distinct insertase function that resides in its TMD and that can be genetically separated from the function of its cytosolic domains.

We hypothesized that Get1/2 interacts with substrate TMDs to couple the release of tail-anchored proteins from Get3 to the subsequent membrane insertion step. To test this idea, we needed a means to stabilize potential Get1/2–substrate interactions during the rapid insertion process. Our solution was to fuse an S-tag to the C terminus of Sec22 and add S protein, which binds to the S-tag to form a stable tertiary structure¹⁶: S protein selectively blocked insertion of full-length Sec22 with a C-terminal S peptide (Sec22s) (Fig. 3a) without compromising the ability of Get1/2 to drive release of the road-blocked substrate from Get3 (Fig. 3b). To establish if road-blocked Sec22s forms a defined complex with Get1/2, we monitored by chemical crosslinking the proximity of an engineered single cysteine in the Sec22 TMD (S192C) and single cysteines at a variety of Get1/2 transmembrane positions that do not perturb Get1/2 function *in vivo* (Extended Data Fig. 7a). When we used bis-maleimidoethane (BMH), which irreversibly crosslinks reactive thiols that are within 13 Å of one another, we detected crosslinks between Sec22 and several positions that lie along the same face of the Get1 TM1 α -helix (Fig. 3c and Extended Data Fig. 7b). We also detected several crosslinks between Sec22 and positions on Get1 TM2 and TM3, as well as a substrate-crosslinking position on Get2 TM3 (Extended Data Fig. 7c). Our data suggest that the road-blocked Sec22 interacts with a composite TMD-docking site that resides near the cytosolic side of the ER membrane, comprises transmembrane regions from both Get1 and

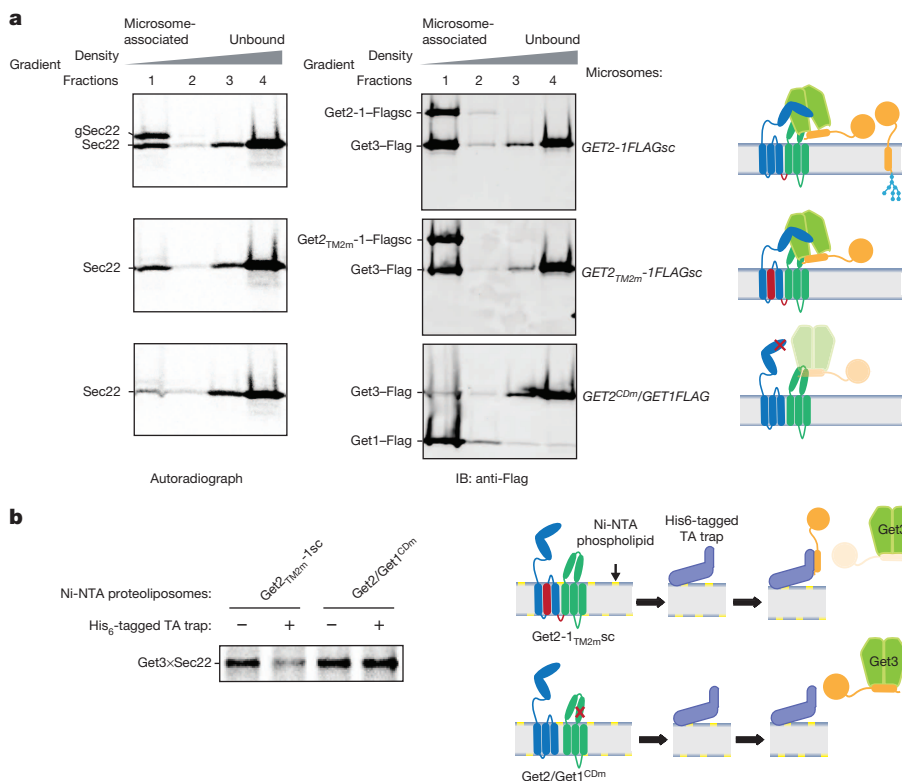


Figure 2 | Get1/2 has a novel insertase function that resides in its TMD. **a**, Affinity-purified Get3–Flag–Sec22 was incubated with the indicated microsomes at room temperature. Samples were overlaid with an Optiprep gradient and subjected to ultracentrifugation. Proteins precipitated from the four indicated gradient fractions were resolved by SDS-PAGE and visualized by autoradiography and immunoblotting. Non-glycosylated Sec22 that associated with GET2-1FLAGsc, GET2^{TM2m}-1FLAGsc and GET2^{CDm}/GET1FLAG microsomes corresponded to 11%, 10% and 2% of the summed Sec22 signal in all the fractions, respectively. **b**, Ni-NTA proteoliposomes reconstituted with the indicated mutant Get1/2 transmembrane complexes were incubated with tail-anchored (TA) trap or mock-incubated followed by crosslinking analysis as in Fig. 1c.

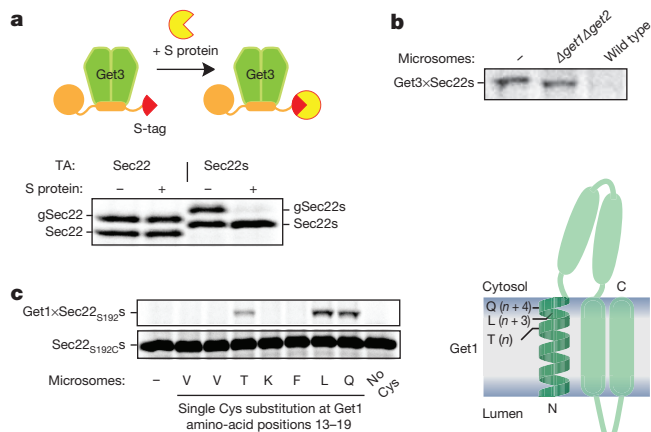


Figure 3 | The Get1/2 TMD binds a road-blocked tail-anchored protein released from Get3. **a**, Top: schematic showing attachment of S protein to the S-tag at the C terminus of a tail-anchored protein bound to Get3. Bottom: affinity-purified Get3–Sec22 or Get3–Sec22 (with S-tag) were incubated with S protein at room temperature or mock-incubated followed by Sec22 insertion analysis with wild-type microsomes as in Fig. 1b. **b**, Affinity-purified Get3–Sec22s was road-blocked with S protein and incubated with the indicated microsomes or mock-incubated at room temperature followed by crosslinking analysis as in Fig. 1c. **c**, Left: affinity-purified Get3–Sec22_{S192C} was road-blocked with S protein and incubated with the indicated microsomes or mock-incubated at room temperature followed by BMH crosslinking on ice. Samples were subjected to SDS–PAGE and visualized by autoradiography. Uncrosslinked Sec22s and an approximately 60 kDa crosslink product between Get1 and Sec22s are indicated. Right: schematic showing the membrane topology of Get1 with crosslinking positions on TM1 indicated.

Get2, and is aqueous in nature, since BMH only reacts with thiols exposed to an aqueous environment¹⁷.

Three lines of evidence argue that by road-blocking Sec22 we have stabilized a bona fide on-pathway (pre-integrated) intermediate. First, we detected no TMD crosslinking to an insertase-defective Get1/2 (Extended Data Fig. 8a, b). Second, as would be predicted for an arrested on-pathway intermediate, addition of S peptide after intermediate accumulation efficiently relieved arrest and resulted in efficient ‘chasing’ to the inserted product (Fig. 4a). Lastly, when we monitored the kinetics of Sec22 TMD crosslinking to two positions on Get1 TM1 during unimpeded integration, we observed the expected rapid and transient intermediate formation (Fig. 4b and Extended Data Fig. 8c). Importantly, Sbh1 (yeast Sec61-β homologue) followed a similar integration pathway to Sec22 (Extended Data Fig. 8d, e), arguing that we have uncovered a general mechanism for insertion of tail-anchored proteins into the ER.

Our work reveals that the Get1/2 transmembrane complex is an insertase that, compared with the Sec61 protein translocon in the ER membrane, provides an essential mechanistic insight into membrane protein insertion in general. Specifically, in the case of co-translational ER membrane protein insertion, the precise mechanism by which the signal recognition particle hands over its substrate to a binding site located within the Sec61 channel is not known but it involves GTP hydrolysis of the signal recognition particle and competitive binding between Sec61 and the signal recognition particle to their overlapping sites near the ribosome exit tunnel^{18,19}. In the case of post-translational tail-anchored protein insertion into the ER, we have shown that Get1/2 cytosolic domains collaborate with the Get1/2 TMD to enable efficient substrate release from Get3. Speculatively, we suggest that substrate handover between Get3 and Get1/2 creates a facilitated pathway that overcomes the kinetic energy barrier between substrate release from the open conformation of Get3 and the thermodynamically favourable substrate state in

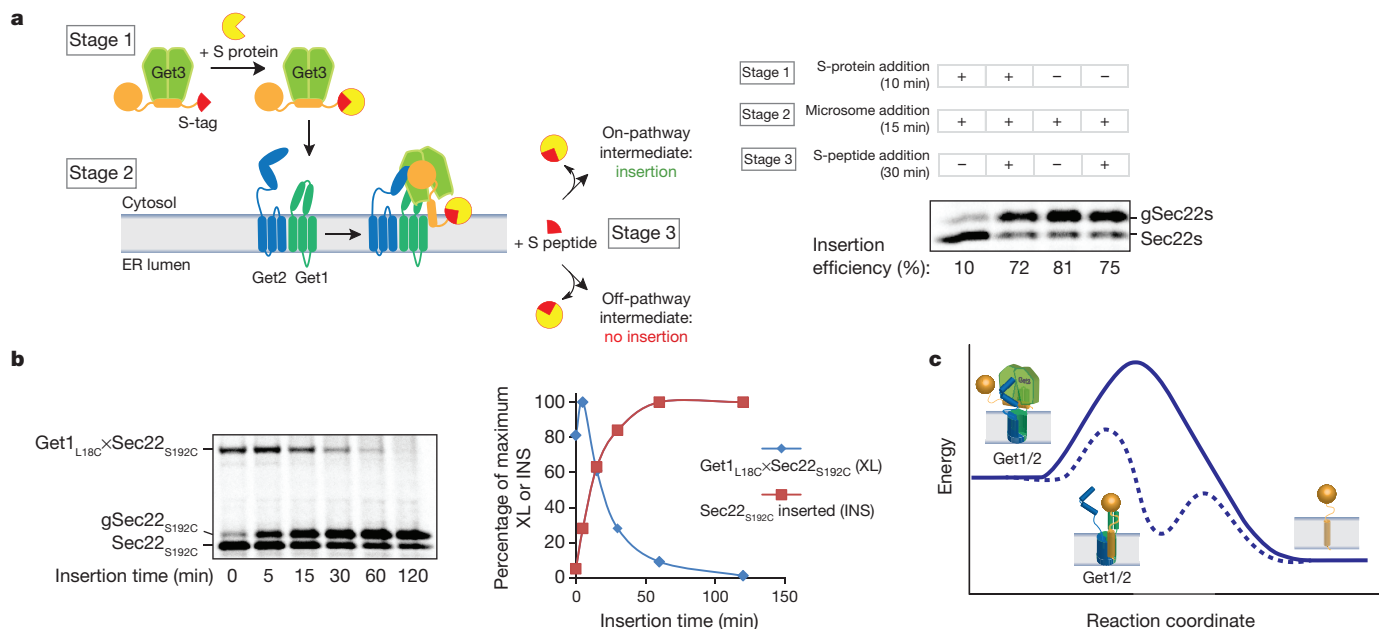


Figure 4 | Tail-anchored protein docking to the Get1/2 TMD is en route a facilitated integration pathway. **a**, Left: schematic illustrating two extreme outcomes of the S peptide ‘chase’ experiment. Right: affinity-purified Get3–Sec22s was split four-ways and subjected to the three stages of the experiment, as indicated, followed by Sec22s insertion analysis with wild-type microsomes as in Fig. 1b. **b**, Affinity-purified Get3–Sec22_{S192C} was incubated with Get1_{L18C} microsomes at room temperature for the indicated amounts of time. Samples were simultaneously subjected to Sec22 insertion and crosslinking analysis as in Figs 1b and 3c, respectively. On the right is a kinetic analysis plot of the gel data shown on the left. XL refers to the crosslinked product between Get1 and Sec22; INS, Sec22_{S192C} inserted. **c**, Working model

for the Get1/2 insertase mechanism. The cytosolic domains work together with a membrane-embedded TMD-docking site to afford a coupled mechanism that facilitates substrate release from Get3 and TMD partitioning into the phospholipid bilayer. We speculate that substrate binding to the TMD-docking site reduces the activation energy barrier (dotted line) that is apparently associated with spontaneous insertion of substrates delivered to insertase-defective Get1/2 (solid line). One way in which the TMD-docking site might afford an integration path around the hydrophilic headgroup barrier is in the form of an aqueous pocket lined on one side by phospholipid hydrocarbon chains, and on the other by a composite surface of Get1 and Get2 transmembrane helices.

which the TMD has partitioned into the lipid bilayer (Fig. 4c). Thus, the Get1/2 insertase mechanism, stripped away of the complexities associated with insertion coupled to co-translational protein translocation, reveals the fundamental molecular logic of membrane protein targeting and insertion: a mechanism for reducing the ability of a targeting factor to shield its substrates enables substrate handover to a binding site that is membrane-embedded within an insertion machine.

METHODS SUMMARY

Standard methods for monitoring Get1/2 function *in vivo* and *in vitro*, as well as *Saccharomyces cerevisiae* strains and plasmids used in this study, are described in detail in Methods.

For protein engineering, Get1 was linked to Get2 using the sequence ASGAGG SEGGGSEGGTSGAT, previously used to link the subunits of a hexameric AAA⁺ ATPase²⁰. S-protein attachment was achieved by mixing 2 µl of the appropriate affinity-purified Get3 targeting complexes with 0.4 µl of S-protein (EMD Millipore) in 8 µl of insertion buffer (22 mM HEPES-KOH pH 7.4, 1.5 mM Mg(OAc)₂, 120 mM KOAc, 2 mM DTT, 14% glycerol, 0.75 mM ATP, 25 mM creatine phosphate and 330 µg ml⁻¹ creatine kinase) for 10 min at room temperature. In the experiment shown in Fig. 4a, the S-protein road-block was alleviated by the addition of synthetic S peptide (1 µg µl⁻¹) (KETAAAKFERQHMDs; Eton Bioscience).

To release tail-anchored substrate from Get3, 2 µl of affinity-purified Get3 targeting complex was mixed with 1 µl of microsomes (attenuance $D_{280\text{ nm}} = 80$) in 8 µl of insertion buffer with ATP regenerating system and incubated for 30 min at room temperature (23 °C). This was followed by crosslinking with 0.5 mM disuccinimidyl suberate (Pierce) at room temperature for 30 min. Crosslinking was quenched by incubation with 50 mM Tris-HCl, pH 7.4 for 5 min at room temperature. Before monitoring Sec22 release from Get3 by proteoliposomes (Fig. 2a), we added 0.3 µg of tail-anchored trap to 1 µl of proteoliposomes and incubated for 20 min at room temperature. In one instance (Extended Data Fig. 1), membranes were substituted with miniGet1/2 and tail-anchored trap (0.1 µg µl⁻¹) or mock trap (0.1 µg µl⁻¹), as described previously⁸.

For TMD docking to Get1/2, 2 µl of affinity-purified Get3 targeting complex was mixed with 1 µl of microsomes ($D_{280\text{ nm}} = 80$) in 8 µl of insertion buffer at room temperature followed by crosslinking with 0.2 mM BMH (Pierce) on ice for 60 min. Crosslinking was quenched by incubation with 50 mM DTT for 15 min on ice.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2013; accepted 13 May 2014.

Published online 20 July 2014.

- Denic, V., Dötsch, V. & Sinning, I. Endoplasmic reticulum targeting and insertion of tail-anchored membrane proteins by the GET pathway. *Cold Spring Harb. Perspect. Biol.* **5**, a013334 (2013).
- Denic, V. A portrait of the GET pathway as a surprisingly complicated young man. *Trends Biochem. Sci.* **37**, 411–417 (2012).
- Hegde, R. S. & Keenan, R. J. Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nature Rev. Mol. Cell Biol.* **12**, 787–798 (2011).

- Chartron, J. W., Clemons, W. M. Jr & Suloway, C. J. M. The complex process of GETting tail-anchored membrane proteins to the ER. *Curr. Opin. Struct. Biol.* **22**, 217–224 (2012).
- Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
- Schuldiner, M. *et al.* The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645 (2008).
- Favaloro, V., Spasic, M., Schwappach, B. & Dobberstein, B. Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* **121**, 1832–1840 (2008).
- Wang, F., Whynot, A., Tung, M. & Denic, V. The mechanism of tail-anchored protein insertion into the ER membrane. *Mol. Cell* **43**, 738–750 (2011).
- Stefer, S. *et al.* Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**, 758–762 (2011).
- Mariappan, M. *et al.* The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–66 (2011).
- Kubota, K., Yamagata, A., Sato, Y., Goto-Ito, S. & Fukui, S. Get1 stabilizes an open dimer conformation of get3 ATPase by binding two distinct interfaces. *J. Mol. Biol.* **422**, 366–375 (2012).
- Borgese, N. & Fasana, E. Targeting pathways of C-tail-anchored proteins. *Biochim. Biophys. Acta* **1808**, 937–946 (2011).
- Leznicki, P., Warwicker, J. & High, S. A biochemical analysis of the constraints of tail-anchored protein biogenesis. *Biochem. J.* **436**, 719–727 (2011).
- Wang, F., Brown, E. C., Mak, G., Zhuang, J. & Denic, V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**, 159–171 (2010).
- Brandman, O. *et al.* A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
- Richards, F. M. & Vithayathil, P. J. The preparation of subtilisin-modified ribonuclease and the separation of the peptide and protein components. *J. Biol. Chem.* **234**, 1459–1465 (1959).
- Li, J. *et al.* Reactions of cysteines substituted in the amphipathic N-terminal tail of a bacterial potassium channel with hydrophilic and hydrophobic maleimides. *Proc. Natl Acad. Sci. USA* **99**, 11605–11610 (2002).
- Akopian, D., Shen, K., Zhang, X. & Shan, S. Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.* **82**, 693–721 (2013).
- Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu. Rev. Biophys.* **41**, 21–40 (2012).
- Martin, A., Baker, T. A. & Sauer, R. T. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**, 1115–1120 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Brandman and R. Hegde for reagents, members of the Denic laboratory for scientific advice, and J. Weissman, A. Murray, B. Stern and C. Patil for comments on the manuscript. This work was supported by the National Institutes of Health (R01GM099994-01) and a postdoctoral fellowship from the Sara Elizabeth O'Brien Trust Postdoctoral Fellowship Program, Bank of America, Co-Trustee (to F.W.).

Author Contributions F.W. performed most of the experiments described in the study. C.C. analyzed Get1 cysteine N-ethylmaleimide (NEM) accessibility. N.R.W. performed cell microscopy experiments. F.W., C.C., N.R.W. and V.D. examined the data. V.D. conceived the project, guided the experiments, and wrote the paper with F.W. and input from C.C. and N.R.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.D. (vdenic@mcb.harvard.edu).

Comparative analysis of the transcriptome across distant species

Mark B. Gerstein^{1,2,3*}, Joel Rozowsky^{1,2*}, Koon-Kiu Yan^{1,2*}, Daifeng Wang^{1,2*}, Chao Cheng^{4,5*}, James B. Brown^{6,7*}, Carrie A. Davis^{8*}, LaDeana Hillier^{9*}, Cristina Sisu^{1,2*}, Jingyi Jessica Li^{7,10,11*}, Baikang Pei^{1,2*}, Arif O. Harmanci^{1,2*}, Michael O. Duff^{1,2*}, Sarah Djebali^{13,14*}, Roger P. Alexander^{1,2}, Burak H. Alver¹⁵, Raymond Auerbach^{1,2}, Kimberly Bell⁸, Peter J. Bickel⁷, Max E. Boeck⁹, Nathan P. Boley^{6,16}, Benjamin W. Booth⁶, Lucy Cherbas^{17,18}, Peter Cherbas^{17,18}, Chao Di¹⁹, Alex Dobin⁸, Jorg Drenkow⁸, Brent Ewing⁹, Gang Fang^{1,2}, Megan Fastuca⁸, Elise A. Feingold²⁰, Adam Frankish²¹, Guanjin Gao¹⁹, Peter J. Good²⁰, Roderic Guigo^{13,14}, Ann Hammonds⁶, Jen Harrow²¹, Roger A. Hoskins⁶, Cédric Howald^{22,23}, Long Hu¹⁹, Haiyan Huang⁷, Tim J. P. Hubbard^{21,24}, Chau Huynh⁹, Sonali Jha⁸, Dionna Kasper²⁵, Masaomi Kato²⁶, Thomas C. Kaufman¹⁷, Robert R. Kitchen^{1,2}, Erik Ladewig²⁷, Julien Lagarde^{13,14}, Eric Lai²⁷, Jing Leng^{1,2}, Zhi Lu¹⁹, Michael MacCoss⁹, Gemma May^{12,28}, Rebecca McWhirter²⁹, Gennifer Merrihew⁹, David M. Miller²⁹, Ali Mortazavi^{30,31}, Rabi Murad^{30,31}, Brian Oliver³², Sara Olson¹², Peter J. Park¹⁵, Michael J. Pazin²⁰, Norbert Perrimon^{33,34}, Dmitri Pervouchine^{13,14}, Valerie Reinke²⁵, Alexandre Reymond²², Garrett Robinson⁷, Anastasia Samsonova^{33,34}, Gary I. Saunders^{21,35}, Felix Schlesinger⁸, Anurag Sethi^{1,2}, Frank J. Slack²⁶, William C. Spencer²⁹, Marcus H. Stoiber^{6,16}, Pnina Strassburger⁹, Andrea Tanzer^{36,37}, Owen A. Thompson⁹, Kenneth H. Wan⁶, Guilin Wang²⁵, Huaian Wang⁸, Kathie L. Watkins²⁹, Jiayu Wen²⁷, Kejia Wen¹⁹, Chenghai Xue⁸, Li Yang^{12,38}, Kevin Yip^{39,40}, Chris Zaleski⁸, Yan Zhang^{1,2}, Henry Zheng^{1,2}, Steven E. Brenner^{41,42}, Brenton R. Graveley¹², Susan E. Celniker⁶, Thomas R. Gingeras⁸ & Robert Waterston⁹

The transcriptome is the readout of the genome. Identifying common features in it across distant species can reveal fundamental principles. To this end, the ENCODE and modENCODE consortia have generated large amounts of matched RNA-sequencing data for human, worm and fly. Uniform processing and comprehensive annotation of these data allow comparison across metazoan phyla, extending beyond earlier within-phylum transcriptome comparisons and revealing ancient, conserved features^{1–6}. Specifically, we discover co-expression modules shared across animals, many of which are enriched in developmental genes. Moreover, we use expression patterns to align the stages in worm and fly development and find a novel pairing between worm embryo and fly pupae, in addition to the embryo-to-embryo and larvae-to-larvae pairings. Furthermore, we find that the extent of non-canonical, non-coding transcription is similar in each organism, per base pair. Finally, we find in all three organisms that the gene-expression levels, both coding and non-coding, can be quantitatively predicted from

chromatin features at the promoter using a ‘universal model’ based on a single set of organism-independent parameters.

Our comparison used the ENCODE–modENCODE RNA resource (Extended Data Fig. 1). This resource comprises: deeply sequenced RNA-sequencing (RNA-seq) data from many distinct samples from all three organisms; comprehensive annotation of transcribed elements; and uniformly processed, standardized analysis files, focusing on non-coding transcription and expression patterns. Where practical, these data sets match comparable samples across organisms and to other types of functional genomics data. In total, the resource contains 575 different experiments containing >67 billion sequence reads. It encompasses many different RNA types, including poly(A)+, poly(A)-, ribosomal-RNA-depleted, short and long RNA.

The annotation in the resource represents a capstone for the decade-long efforts in human, worm and fly. The new annotation sets have numbers, sizes and families of protein-coding genes similar to previous

¹Program in Computational Biology and Bioinformatics, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. ²Department of Molecular Biophysics and Biochemistry, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. ³Department of Computer Science, Yale University, 51 Prospect Street, New Haven, Connecticut 06511, USA. ⁴Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755, USA. ⁵Institute for Quantitative Biomedical Sciences, Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire 03766, USA. ⁶Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁷Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley, California 94720-3860, USA. ⁸Functional Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ⁹Department of Genome Sciences and University of Washington School of Medicine, William H. Foege Building S350D, 1705 Northeast Pacific Street, Box 355065 Seattle, Washington 98195-0665, USA. ¹⁰Department of Statistics, University of California, Los Angeles, California 90095-1554, USA. ¹¹Department of Human Genetics, University of California, Los Angeles, California 90095-7088, USA. ¹²Department of Genetics and Developmental Biology, Institute for Systems Genomics, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, Connecticut 06030, USA. ¹³Centre for Genomic Regulation, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain. ¹⁴Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Catalonia, Spain. ¹⁵Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, Massachusetts 02115, USA. ¹⁶Department of Biostatistics, University of California, Berkeley, 367 Evans Hall, Berkeley, California 94720-3860, USA. ¹⁷Department of Biology, Indiana University, 1001 East 3rd Street, Bloomington, Indiana 47405-7005, USA. ¹⁸Center for Genomics and Bioinformatics, Indiana University, 1001 East 3rd Street, Bloomington, Indiana 47405-7005, USA. ¹⁹MOE Key Lab of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²⁰National Human Genome Research Institute, National Institutes of Health, 5635 Fishers Lane, Bethesda, Maryland 20892-9307, USA. ²¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²²Center for Integrative Genomics, University of Lausanne, Genopode building, Lausanne 1015, Switzerland. ²³Swiss Institute of Bioinformatics, Genopode building, Lausanne 1015, Switzerland. ²⁴Medical and Molecular Genetics, King's College London, London WC2R 2LS, UK. ²⁵Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520-8005, USA. ²⁶Department of Molecular, Cellular and Developmental Biology, PO Box 208103, Yale University, New Haven, Connecticut 06520, USA. ²⁷Sloan-Kettering Institute, 1275 York Avenue, Box 252, New York, New York 10065, USA. ²⁸Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 USA. ²⁹Department of Cell and Developmental Biology, Vanderbilt University, 465 21st Avenue South, Nashville, Tennessee 37232-8240, USA. ³⁰Developmental and Cell Biology, University of California, Irvine, California 92697, USA. ³¹Center for Complex Biological Systems, University of California, Irvine, California 92697, USA. ³²Section of Developmental Genomics, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ³³Department of Genetics and Drosophila RNAi Screening Center, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ³⁴Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ³⁵European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK. ³⁶Bioinformatics and Genomics Programme, Center for Genomic Regulation, Universitat Pompeu Fabra (CRG-UPF), 08003 Barcelona, Catalonia, Spain. ³⁷Institute for Theoretical Chemistry, Theoretical Biochemistry Group (TBI), University of Vienna, Währingerstrasse 17/3/303, A-1090 Vienna, Austria. ³⁸Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. ³⁹Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. ⁴⁰CUHK-BGI Innovation Institute of Trans-omics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. ⁴¹Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ⁴²Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

†These authors jointly supervised this work.

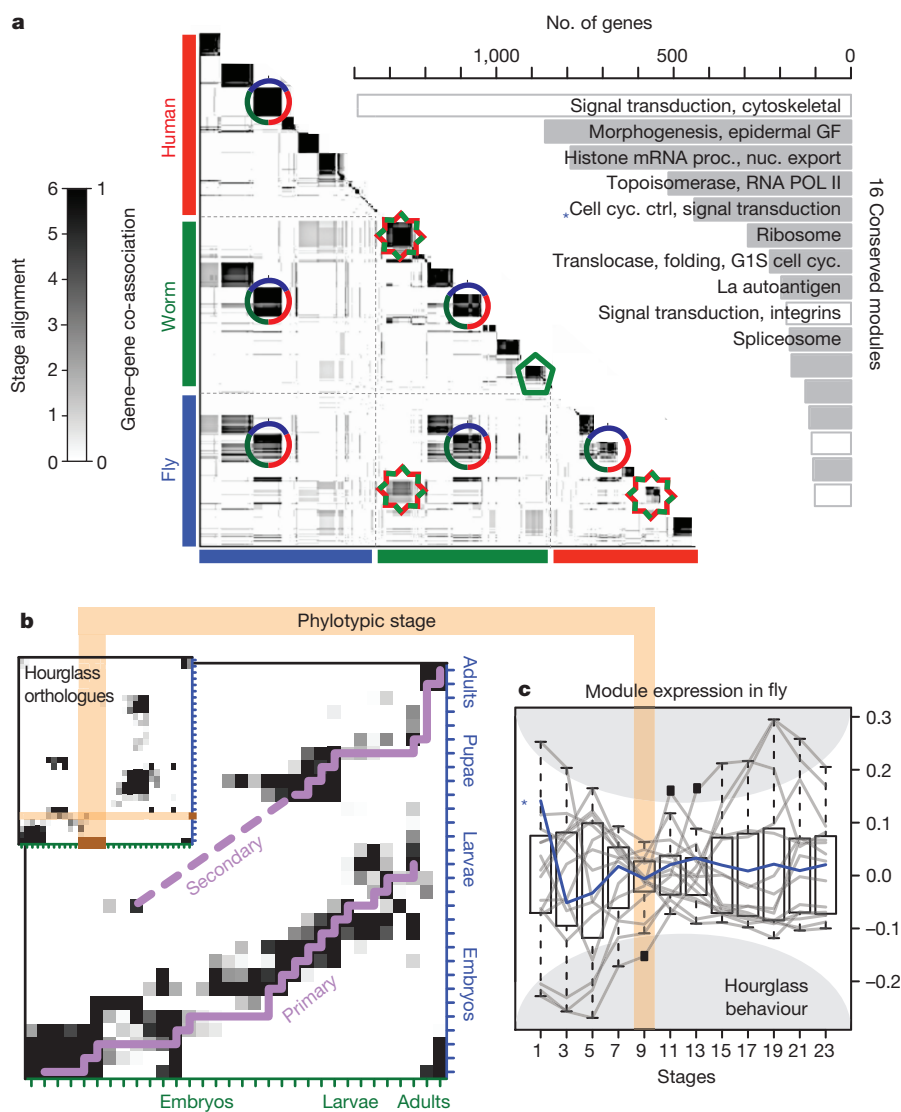


Figure 1 | Expression clustering. **a**, Left, human, worm and fly gene-gene co-association matrix; darker colouring reflects the increased likelihood that a pair of genes are assigned to the same module. A dark block along the diagonal represents a group of genes within a species. If this is associated with an off-diagonal block then it is a cross-species module (for example, a three-species conserved module is shown with a circle and a worm-fly module, with a star). However, if a diagonal block has no off-diagonal associations, then it forms a species-specific module (for example, green pentagon). Right, the Gene Ontology functional enrichment of genes within the 16 conserved modules is shown. GF, growth factor; nuc., nuclear; proc., processing. **b**, Primary and secondary alignments of worm-and-fly developmental stages based on all worm-fly orthologues. Inset shows worm-fly stage alignment using only hourglass orthologues is more significant and exhibits a gap (brown) matching the phylotypic stage. The scale for the heat map in **b** is indicated on the left side of the scale in **a** (labelled stage alignment). **c**, Normalized expression of the conserved modules in fly shows the smallest intra-organism divergence during the phylotypic stage (brown). A representative module is indicated with a blue asterisk in **a** and **c**. (For further details see Extended Data Figs 5 and 6; ref. 20, related to the left part of **a**; and ref. 21, related to the bottom part of **b**.)

compilations; however, the number of pseudogenes and annotated non-coding RNAs differ (Extended Data Fig. 2, Extended Data Table 1 and Supplementary Fig. 1). Also, the number of splicing events is greatly increased, resulting in a concomitant increase in protein complexity. We find the proportion of the different types of alternative splicing (for example, exon skipping or intron retention) is generally similar across the three organisms; however, skipped exons predominate in human while retained introns are most common in worm and fly⁷ (Extended Data Fig. 3, Supplementary Fig. 1 and Supplementary Table 1).

A fraction of the transcription comes from genomic regions not associated with standard annotations, representing 'non-canonical transcription' (Supplementary Table 2)⁸. Using a minimum-run-maximum-gap algorithm to process reads mapping outside of protein-coding transcripts, pseudogenes and annotated non-coding RNAs, we identified read clusters; that is, transcriptionally active regions (TARs). Across all three genomes we found roughly one-third of the bases gives rise to TARs or non-canonical transcription (Extended Data Table 1). To determine the extent that this transcription represents an expansion of the current established classes of non-coding RNAs, we identified the TARs most similar to known annotated non-coding RNAs using a supervised classifier⁹ (Supplementary Fig. 2 and Supplementary Table 2). We validated the classifier's predictions using RT-PCR (PCR with reverse transcription), demonstrating high accuracy. Overall, these predictions encompass only a small fraction of all TARs, suggesting that most TARs have features distinct from annotated non-coding RNAs and that the majority of

non-coding RNAs of established classes have already been identified. To shed further light on the possible roles of TARs we intersected them with enhancers and HOT (high-occupancy target) regions^{8,10-13}, finding statistically significant overlaps (Extended Data Fig. 4 and Supplementary Table 2).

Given the uniformly processed nature of the data and annotations, we were able to make comparisons across organisms. First, we built co-expression modules, extending earlier analysis¹⁴ (Fig. 1a). To detect modules consistently across the three species, we combined across-species orthology and within-species co-expression relationships. In the resulting multilayer network we searched for dense subgraphs (modules), using simulated annealing^{15,16}. We found some modules dominated by a single species, whereas others contain genes from two or three. As expected, the modules with genes from multiple species are enriched in orthologues. Moreover, a phylogenetic analysis shows that the genes in such modules are more conserved across 56 diverse animal species (Extended Data Fig. 5 and Supplementary Fig. 3). To focus on the cross-species conserved functions, we restricted the clustering to orthologues, arriving at 16 conserved modules, which are enriched in a variety of functions, ranging from morphogenesis to chromatin remodelling (Fig. 1a and Supplementary Table 3). Finally, we annotated many TARs based on correlating their expression profiles with these modules (Extended Data Fig. 4).

Next, we used expression profiles of orthologous genes to align the developmental stages in worm and fly (Fig. 1b and Extended Data Fig. 6). For every developmental stage, we identified stage-associated genes; that

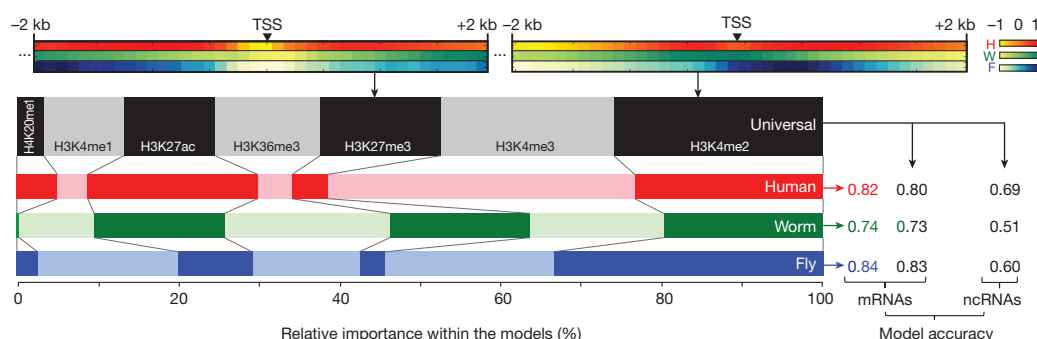


Figure 2 | Histone models for gene expression. Top, normalized correlations of two representative histone marks with expression. Left, relative importance of the histone marks in organism-specific models and the universal model. Right, prediction accuracies (Pearson correlations all significant, $P < 1 \times 10^{-100}$) of the organism-specific and universal models. (See Extended Data Figs 7 and 8 for further details.)

is, genes highly expressed at that particular stage but not across all stages. We then counted the number of orthologous pairs among these stage-associated genes for each possible worm-and-fly stage correspondence, aligning stages by the significance of the overlap. Notably, worm stages map to two sets of fly stages. First, they match in a co-linear fashion to the fly (that is, embryos-to-embryos, larvae-to-larvae). However, worm late embryonic stages also match fly pupal stages, suggesting a shared expression program between embryogenesis and metamorphosis. The approximately 50 stage-associated genes involved in this dual alignment are enriched in functions such as ion transport and cation-channel activity (Supplementary Table 3).

To gain further insight into the stage alignment, we examined our 16 conserved modules in terms of the 'hourglass hypothesis', which posits that all animals go through a particular stage in embryonic development (the tight point of the hourglass or 'phylotypic' stage) during which the expression divergence across species for orthologous genes is smallest^{4,5,17}. For genes in 12 of the 16 modules, we observed canonical hourglass behaviour; that is, inter-organism expression divergence across closely related fly species during development is minimal⁵ (Supplementary Fig. 3). Moreover, we find a subset of TARs also exhibit this hourglass behaviour (Supplementary Fig. 2). Beyond looking at inter-species divergence, we also investigated the *intra*-species divergence within just *Drosophila melanogaster* and *Caenorhabditis elegans*. Notably, we observed that divergence of gene expression between modules is minimized during the worm and fly phylotypic stages (Fig. 1c). This suggests, for an individual species, the expression patterns of different modules are most tightly coordinated (low divergence) during the phylotypic stage, but each module has its own expression signature before and after this. In fact it is possible to see this coordination directly as a local maximum in between-module correlations for the worm (Extended Data Fig. 5). Finally, using genes from just the 12 'hourglass modules', we found that the alignment between worm and fly stages becomes stronger (Fig. 1b and Supplementary Fig. 3); in particular it shows a gap where no changes are observed, perfectly matching the phylotypic stage.

The uniformly processed and matched nature of the transcriptome data also facilitates integration with upstream factor-binding and chromatin-modification signals. We investigated the degree to which these upstream signals can quantitatively predict gene expression and how consistent this prediction is across organisms. Similar to previous reports^{11,18,19}, we found consistent correlations, around the transcription start site (TSS), in each of the three species between various histone-modification signals and the expression level of the downstream gene: H3K4me1, H3K4me2, H3K4me3 and H3K27ac are positively correlated, whereas H3K27me3 is negatively correlated (Fig. 2, Extended Data Fig. 7 and Supplementary Fig. 4). Then for each organism, we integrated these individual correlations into a multivariate, statistical model, obtaining high accuracy in predicting expression for protein-coding genes and non-coding RNAs. The promoter-associated marks, H3K4me2 and H3K4me3, consistently have the highest contribution to the model.

A similar statistical analysis with transcription factors showed the correlation between gene expression and transcription-factor binding to be the greatest at the TSS, positively for activators and negatively for repressors (Extended Data Fig. 7). Integrated transcription-factor models in

each organism also achieved high accuracy for protein-coding genes and non-coding RNAs, with as few as five transcription factors necessary for accurate predictions (Extended Data Fig. 8). This perhaps reflects an intricate, correlated structure to regulation. The relative importance of the upstream regions is more peaked for the transcription-factor models than for the histone ones, likely reflecting the fact that histone modifications are spread over broader regions, including the gene body, whereas most transcription factors bind near the promoter.

Finally, we constructed a 'universal model', containing a single set of organism-independent parameters (Fig. 2 and Supplementary Fig. 4). This achieved accuracy comparable to the organism-specific models. In the universal model, the consistently important promoter-associated marks such as H3K4me2 and H3K4me3 are weighted most highly. In contrast, the enhancer mark H3K4me1 is down-weighted, perhaps reflecting that signals for most human enhancers are not near the TSS. Using the same set of organism-independent parameters derived from training on protein-coding genes, the universal model can also accurately predict non-coding RNA expression.

Overall, our comparison of the transcriptomes of three phylogenetically distant metazoans highlights fundamental features of transcription conserved across animal phyla. First, there are ancient co-expression modules across organisms, many of which are enriched for developmentally important hourglass genes. These conserved modules have highly coordinated intra-organism expression during the phylotypic stage, but display diversified expression before and after. The expression clustering also aligns developmental stages between worm and fly, revealing shared expression programs between embryogenesis and metamorphosis. Finally, we were able to build a single model that could predict transcription in all three organisms from upstream histone marks using a single set of parameters for both protein-coding genes and non-coding RNAs. Overall, our results underscore the importance of comparing divergent model organisms to human to highlight conserved biological principles (and disentangle them from lineage-specific adaptations).

METHODS SUMMARY

Detailed methods are given in the Supplementary Information. (See the first section of the Supplementary Information for a guide.) More details on data availability are given in section F of the Supplementary Information.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 April 2013; accepted 30 April 2014.

1. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
2. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**, 1593–1599 (2012).
3. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
4. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* **22**, 1101–1108 (2012).
5. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).

6. Simola, D.F., Francis, C., Sniegowski, P.D. & Kim, J. Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biol.* **11**, R105 (2010).
7. Talerico, M. & Berget, S.M. Intron definition in splicing of small *Drosophila* introns. *Mol. Cell. Biol.* **14**, 3434–3445 (1994).
8. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
9. Lu, Z.J. *et al.* Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res.* **21**, 276–285 (2011).
10. Boyle, A.P. *et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* <http://dx.doi.org/10.1038/nature13668> (this issue).
11. Gerstein, M.B. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
12. modENCODE Consortium, *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1798–1797 (2010).
13. Ho, J.W.K. *et al.* modENCODE and ENCODE resources for analysis of metazoan chromatin organization. *Nature* <http://dx.doi.org/10.1038/nature13497> (this issue).
14. Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
15. Kirkpatrick, S., Gelatt, C.D., Jr & Vecchi, M.P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
16. Reichardt, J. & Bornholdt, S. Detecting fuzzy community structures in complex networks with a Potts model. *Phys. Rev. Lett.* **93**, 218701 (2004).
17. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
18. Karlič, R., Chung, H.-R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc. Natl Acad. Sci. USA* **107**, 2926–2931 (2010).
19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
20. Yan, K.K., Wang, D., Rozowsky, J., Zheng, H., Cheng, C. & Gerstein, M. OrthoClust: an orthology-based network framework for clustering data across multiple species. *Genome Biol.* **15**, R100 (2014).
21. Li, J.J., Huang, H., Bickel, P.J. & Brenner, S.E. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res.* **24**, 1086–1101 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank the NHGRI and the ENCODE and modENCODE projects for support. In particular, this work was funded by a contract from the National Human Genome Research Institute modENCODE Project, contract U01 HG004271 and U54 HG006944, to S.E.C. (principal investigator) and P.C., T.R.G., R.A.H. and B.R.G. (co-principal investigators) with additional support from R01 GM076655 (S.E.C.) both

under Department of Energy contract no. DE-AC02-05CH11231, and U54 HG007005 to B.R.G. J.B.B.'s work was supported by NHGRI K99 HG006698 and DOE DE-AC02-05CH11231. Work in P.J.B.'s group was supported by the modENCODE DAC sub award 5710003102, 1U01HG007031-01 and the ENCODE DAC 5U01HG004695-04. Work in M.B.G.'s group was supported by NIH grants HG007000 and HG007355. Work in Bloomington was supported in part by the Indiana METACyt Initiative of Indiana University, funded by an award from the Lilly Endowment, Inc. Work in E.C.L.'s group was supported by U01-HG004261 and RC2-HG005639. P.J.P. acknowledges support from the National Institutes of Health (grant no. U01HG004258). We thank the HAVANA team for providing annotation of the human reference genome, whose work is supported by National Institutes of Health (grant no. 5U54HG004555), the Wellcome Trust (grant no. WT098051). R.G. acknowledges support from the Spanish Ministry of Education (grant BIO2011-26205). We also acknowledge use of the Yale University Biomedical High Performance Computing Center. R.W.'s lab was supported by grant no. U01 HG 004263.

Author Contributions Work on the paper was divided between data production and analysis. The analysts were J.R., K.K.Y., D.W., C.C., J.B.B., C.S., J.J.L., B.P., A.O.H., M.O.D., S.D., R.P.A., B.H.A., R.K.A., P.J.B., N.P.B., C.D., A.D., G.F., A.F., R.G., J.H., L.H., H.H., T.H., R.R.K., J.L., J.L., Z.L., A.M., R.M., P.P., D.P., A.S., K.W., K.Y., Y.Z. and H.Z. (names are sorted according to their order in the author list). The data producers were C.A.D., L.H., K.B., M.E.B., B.W.B., L.C., P.C., J.D., B.E., M.F., G.G., P.G., A.H., R.A.H., C.H., C.H., S.J., D.K., M.K., T.C.K., E.L., E.L., M.M., G.M., R.M., G.M., D.M.M., B.O., S.O., N.P., V.R., A.R., G.R., A.S., G.I.S., F.S., F.J.S., W.C.S., M.H.S., P.S., K.L.W., J.W., C.X., L.Y. and C.Z. Substantially larger contributions were made by the joint first authors. The role of the NIH Project Management Group, E.A.F., P.J.G., M.J.P., was limited to coordination and scientific management of the modENCODE and ENCODE consortia. Overall project management was carried out by the senior authors M.B.G., R.W., T.R.G., S.E.C., B.R.G. and S.E.B.

Author Information Data sets described here can be obtained from the ENCODE project website at <http://www.encodeproject.org/comparative> via accession number ENCSR145VDW (alternate URL <http://cmptxn.gersteinlab.org>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B.G., R.W., T.R.G., S.E.C., B.R.G. or S.E.B. (cmptxn@gersteinlab.org).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Comparative analysis of metazoan chromatin organization

Joshua W. K. Ho^{1,2,*†}, Youngsook L. Jung^{1,2,*}, Tao Liu^{3,4,*†}, Burak H. Alver¹, Soohyun Lee¹, Kohta Ikegami^{5†}, Kyung-Ah Sohn^{6,7}, Aki Minoda^{8,9†}, Michael Y. Tolstorukov^{1,2,10}, Alex Appert¹¹, Stephen C. J. Parker^{12,13}, Tingting Gu¹⁴, Anshul Kundaje^{15,16†}, Nicole C. Riddle^{14†}, Eric Bishop^{1,17}, Thea A. Egelhofer¹⁸, Sheng'en Shawn Hu¹⁹, Artyom A. Alekseyenko^{2,20}, Andreas Rechtsteiner¹⁸, Dalal Asker^{21,22}, Jason A. Belsky²³, Sarah K. Bowman¹⁰, Q. Brent Chen⁵, Ron A.-J. Chen¹¹, Daniel S. Day^{1,24}, Yan Dong¹¹, Andrea C. Dose²⁵, Xikun Duan¹⁹, Charles B. Epstein¹⁶, Sevinc Ercan^{5,26}, Elise A. Feingold¹³, Francesco Ferrari¹⁸, Jacob M. Garrigues¹⁸, Nils Gehlenborg^{1,16}, Peter J. Good¹³, Psalm Haseley^{1,2}, Daniel He⁹, Moritz Herrmann¹¹, Michael M. Hoffman²⁷, Tess E. Jeffers^{5†}, Peter V. Kharchenko¹, Paulina Kolasinska-Zwierz¹¹, Chitra V. Kotwaliwale^{9,28}, Nischay Kumar^{15,16}, Sasha A. Langley^{8,9}, Erica N. Larschan²⁹, Isabel Latorre¹¹, Maxwell W. Libbrecht³⁰, Xueqiu Lin¹⁹, Richard Park^{1,17}, Michael J. Pazin¹³, Hoang N. Pham^{8,9,28}, Annette Plachetka^{2,20}, Bo Qin¹⁹, Yuri B. Schwartz^{21,31}, Noam Shores¹⁶, Przemyslaw Stempor¹¹, Anne Vielle¹¹, Chengyang Wang¹⁹, Christina M. Whittle^{9,28}, Huiling Xue^{1,2}, Robert E. Kingston¹⁰, Ju Han Kim^{7,32}, Bradley E. Bernstein^{16,28,33}, Abby F. Dernburg^{8,9,28}, Vincenzo Pirrotta²¹, Mitzi I. Kuroda^{2,20}, William S. Noble^{30,34}, Thomas D. Tullius^{17,35}, Manolis Kellis^{15,16}, David M. MacAlpine²³, Susan Strome¹⁸, Sarah C. R. Elgin¹⁴, Xiaole Shirley Liu^{3,4,16}, Jason D. Lieb^{5†}, Julie Ahringer¹¹, Gary H. Karpen^{8,9} & Peter J. Park^{1,2,36}

Genome function is dynamically regulated in part by chromatin, which consists of the histones, non-histone proteins and RNA molecules that package DNA. Studies in *Caenorhabditis elegans* and *Drosophila melanogaster* have contributed substantially to our understanding of molecular mechanisms of genome function in humans, and have revealed conservation of chromatin components and mechanisms^{1–3}. Nevertheless, the three organisms have markedly different genome sizes, chromosome architecture and gene organization. On human and fly chromosomes, for example, pericentric heterochromatin flanks single centromeres, whereas worm chromosomes have dispersed heterochromatin-like regions enriched in the distal chromosomal ‘arms’, and centromeres distributed along their lengths^{4,5}. To systematically investigate chromatin organization and associated gene regulation across species, we generated and analysed a large collection of genome-wide chromatin data sets from cell lines and developmental stages in worm, fly and human. Here we present over 800 new data sets from our ENCODE and modENCODE consortia, bringing the total to over 1,400. Comparison of combinatorial patterns of histone modifications, nuclear lamina-associated domains, organization of large-scale topological domains, chromatin environment at promoters and

enhancers, nucleosome positioning, and DNA replication patterns reveals many conserved features of chromatin organization among the three organisms. We also find notable differences in the composition and locations of repressive chromatin. These data sets and analyses provide a rich resource for comparative and species-specific investigations of chromatin composition, organization and function.

We used chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) or microarray hybridization (ChIP-chip) to generate profiles of core histones, histone variants, histone modifications and chromatin-associated proteins (Fig. 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2). Additional data include DNase I hypersensitivity sites in fly and human cells, and nucleosome occupancy maps in all three organisms. Compared to our initial publications^{1–3}, this represents a tripling of available fly and worm data sets and a substantial increase in human data sets (Fig. 1b, c). Uniform quality standards for experimental protocols, antibody validation and data processing were used throughout the projects⁶. Detailed analyses of related transcriptome and transcription factor data are presented in accompanying papers^{7,8}.

We performed systematic cross-species comparisons of chromatin composition and organization, focusing on targets profiled in at least two

¹Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, 450 Brookline Avenue, Boston, Massachusetts 02115, USA. ⁵Department of Biology and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁶Department of Information and Computer Engineering, Ajou University, Suwon 443-749, Korea. ⁷Systems Biomedical Informatics Research Center, College of Medicine, Seoul National University, Seoul 110-799, Korea. ⁸Department of Genome Dynamics, Life Sciences Division, Lawrence Berkeley National Lab, Berkeley, California 94720, USA. ⁹Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. ¹⁰Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ¹¹The Gurdon Institute and Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ¹²National Institute of General Medical Sciences, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹³National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁴Department of Biology, Washington University in St. Louis, St. Louis, Missouri 63130, USA. ¹⁵Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ¹⁶Broad Institute, Cambridge, Massachusetts 02141, USA. ¹⁷Program in Bioinformatics, Boston University, Boston, Massachusetts 02215, USA. ¹⁸Department of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, California 95064, USA. ¹⁹Department of Bioinformatics, School of Life Science and Technology, Tongji University, Shanghai 200092, China. ²⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²¹Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA. ²²Food Science and Technology Department, Faculty of Agriculture, Alexandria University, 21545 El-Shatby, Alexandria, Egypt. ²³Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. ²⁴Harvard/MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA. ²⁵Department of Anatomy Physiology and Cell Biology, University of California Davis, Davis, California 95616, USA. ²⁶Department of Biology, Center for Genomics and Systems Biology, New York University, New York, New York 10003, USA. ²⁷Princess Margaret Cancer Centre, Toronto, Ontario M6G 1L7, Canada. ²⁸Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. ²⁹Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, Rhode Island 02912, USA. ³⁰Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195, USA. ³¹Department of Molecular Biology, Umea University, 901 87 Umea, Sweden. ³²Seoul National University Biomedical Informatics, Division of Biomedical Informatics, College of Medicine, Seoul National University, Seoul 110-799, Korea. ³³Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ³⁴Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ³⁵Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA. ³⁶Informatics Program, Children's Hospital, Boston, Massachusetts 02215, USA. [†]Present addresses: Victor Chang Cardiac Research Institute and The University of New South Wales, Sydney, New South Wales 2052, Australia (J.W.K.H.); Department of Biochemistry, University at Buffalo, Buffalo, New York 14203, USA (T.L.); Department of Molecular Biology and Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08540, USA (K.L., T.E.J.); Department of Human Genetics, University of Chicago, Chicago, Illinois 06037, USA (J.D.L.); Division of Genomic Technologies, Center for Life Science Technologies, RIKEN, Yokohama 230-0045, Japan (A.M.); Department of Genetics, Department of Computer Science, Stanford University, Stanford, California 94305, USA (A.K.); Department of Biology, The University of Alabama at Birmingham, Birmingham, Alabama 35294, USA (N.C.R.).

*These authors contributed equally to this work.

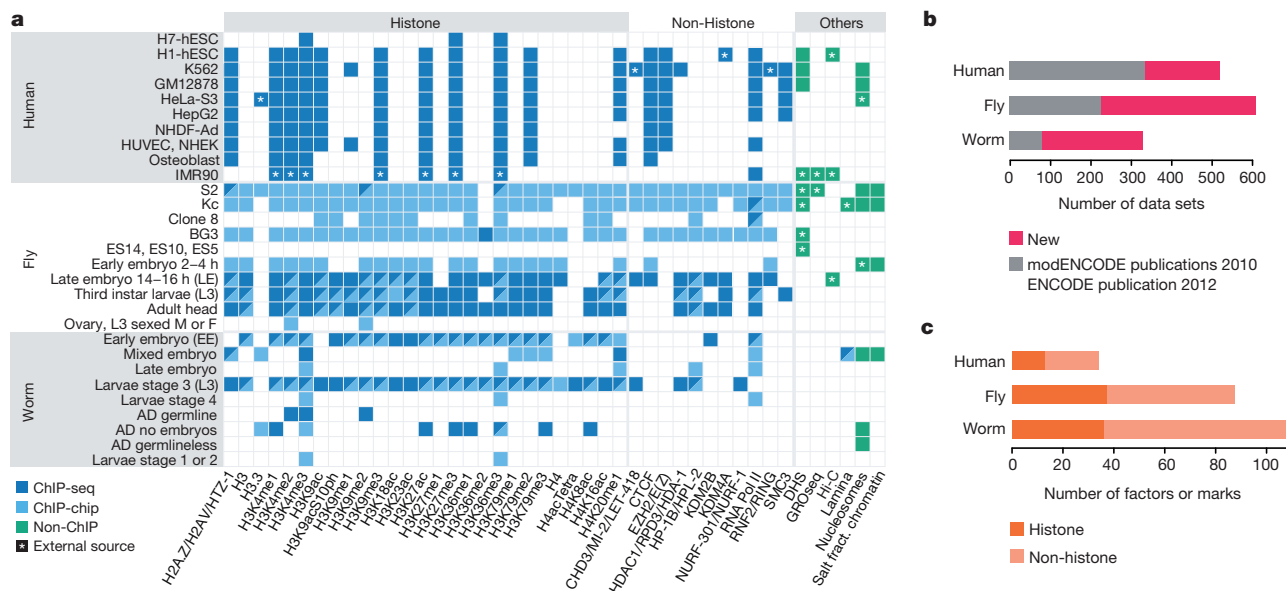


Figure 1 | Overview of the data set. **a**, Histone modifications, chromosomal proteins and other profiles mapped in at least two species (see Supplementary Fig. 1 for the full data set and Supplementary Table 1 for detailed descriptions). Different protein names for orthologues are separated by ‘/’ (see Supplementary Table 2). **b**, The number of all data sets generated by this

organisms (Fig. 1). Sample types used most extensively in our analyses are human cell lines H1-hESC, GM12878 and K562; fly late embryos, third instar larvae and cell lines S2, Kc, BG3; and worm early embryos and stage 3 larvae. Our conclusions are summarized in Extended Data Table 1.

Not surprisingly, the three species show many common chromatin features. Most of the genome in each species is marked by at least one histone modification (Supplementary Fig. 2), and modification patterns are similar around promoters, gene bodies, enhancers and other chromosomal elements (Supplementary Figs 3–12). Nucleosome occupancy patterns around protein-coding genes and enhancers are also largely similar across species, although we observed subtle differences in H3K4me3 enrichment patterns around transcription start sites (TSSs) (Extended Data Fig. 1a and Supplementary Figs 12–14). The configuration and composition of large-scale features such as lamina-associated domains (LADs) are similar (Supplementary Figs 15–17). LADs in human and fly are associated with late replication and H3K27me3 enrichment, suggesting a repressive chromatin environment (Supplementary Fig. 18). Finally, DNA structural features associated with nucleosome positioning are strongly conserved (Supplementary Figs 19 and 20).

Although patterns of histone modifications across active and silent genes are largely similar in all three species, there are some notable differences (Extended Data Fig. 1b). For example, H3K23ac is enriched at promoters of expressed genes in worm, but is enriched across gene bodies of both expressed and silent genes in fly. H4K20me1 is enriched on both expressed and silent genes in human but only on expressed genes in fly and worm (Extended Data Fig. 1b). Enrichment of H3K36me3 in genes expressed with stage or tissue specificity is lower than in genes expressed broadly, possibly because profiling was carried out on mixed tissues (Supplementary Figs 21–23; see Supplementary Methods). Although the co-occurrence of pairs of histone modifications is largely similar across the three species, there are clearly some species-specific patterns (Extended Data Fig. 1c and Supplementary Figs 24 and 25).

Previous studies showed that in human^{9,10} and fly^{1,11} prevalent combinations of marks or ‘chromatin states’ correlate with functional features such as promoters, enhancers, transcribed regions, Polycomb-associated domains, and heterochromatin. ‘Chromatin state maps’ provide a concise and systematic annotation of the genome. To compare chromatin

and previous consortia publications¹⁻³ (new, 815; old, 638). Each data set corresponds to a replicate-merged normalized profile of a histone, histone variant, histone modification, non-histone chromosomal protein, nucleosome or salt-fractionated nucleosome. **c**, The number of unique histone marks or non-histone chromosomal proteins profiled.

states across the three organisms, we developed and applied a novel hierarchical non-parametric machine-learning method called hiHMM (see Supplementary Methods) to generate chromatin state maps from eight histone marks mapped in common, and compared the results with published methods (Fig. 2 and Supplementary Figs 26–28). We find that combinatorial patterns of histone modifications are largely conserved. Based on correlations with functional elements (Supplementary Figs 29–32), we categorized the 16 states into six groups: promoter (state 1), enhancer (states 2 and 3), gene body (states 4–9), Polycomb-repressed (states 10 and 11), heterochromatin (states 12 and 13), and weak or low signal (states 14–16).

Heterochromatin is a classically defined and distinct chromosomal domain with important roles in genome organization, genome stability, chromosome inheritance and gene regulation. It is typically enriched for H3K9me3 (ref. 12), which we used as a proxy for identifying heterochromatic domains (Fig. 3a and Supplementary Figs 33 and 34). As expected, the majority of the H3K9me3-enriched domains in human and fly are concentrated in the pericentromeric regions (as well as other specific domains, such as the Y chromosome and fly 4th chromosome), whereas in worm they are distributed throughout the distal chromosomal ‘arms’^{11,13,14} (Fig. 3a). In all three organisms, we find that more of the genome is associated with H3K9me3 in differentiated cells and tissues compared to embryonic cells and tissues (Extended Data Fig. 2a). We also observe large cell-type-specific blocks of H3K9me3 in human and fly^{11,14,15} (Supplementary Fig. 35). These results suggest a molecular basis for the classical concept of ‘facultative heterochromatin’ formation to silence blocks of genes as cells specialize.

Two distinct types of transcriptionally repressed chromatin have been described. As discussed above, classical ‘heterochromatin’ is generally concentrated in specific chromosomal regions and enriched for H3K9me3 and also H3K9me2 (ref. 12). In contrast, Polycomb-associated silenced domains, involved in cell-type-specific silencing of developmentally regulated genes^{11,14}, are scattered across the genome and enriched for H3K27me3. We found that the organization and composition of these two types of transcriptionally silent domains differ across species. First, human, fly and worm display significant differences in H3K9 methylation patterns. H3K9me2 shows a stronger correlation with H3K9me3 in fly than in worm ($r = 0.89$ versus $r = 0.40$, respectively), whereas H3K9me2

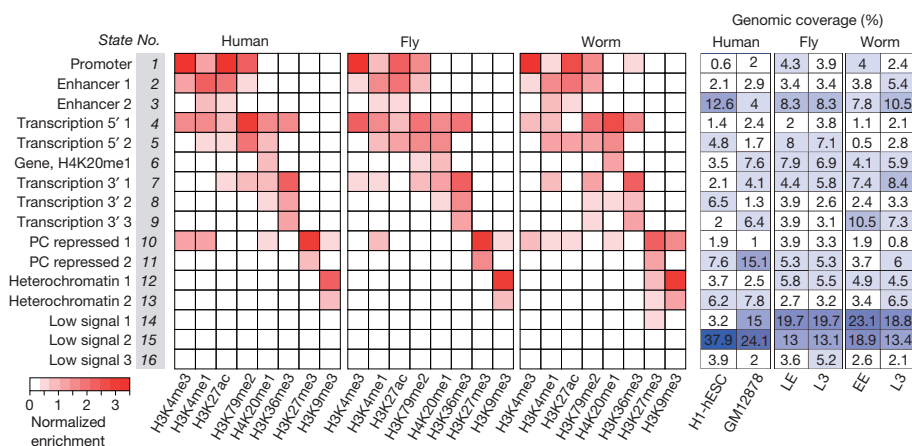


Figure 2 | Shared and organism-specific chromatin states. Sixteen chromatin states derived by joint segmentation using hiHMM (see Supplementary Methods) based on enrichment patterns of eight histone

is well correlated with H3K9me1 in worm but not in fly ($r = 0.44$ versus $r = -0.32$, respectively) (Fig. 3b). These findings suggest potential differences in heterochromatin in the three organisms (see below). Second, the chromatin state maps reveal two distinct types of Polycomb-associated repressed regions: strong H3K27me3 accompanied by marks for active genes or enhancers (Fig. 2, state 10; perhaps due to mixed tissues in whole embryos or larvae for fly and worm), and strong H3K27me3 without active marks (state 11) (see also Supplementary Fig. 31). Third, we observe a worm-specific association of H3K9me3 and H3K27me3. These two marks are enriched together in states 12 and 13 in worm but not in human and fly. This unexpectedly strong association between H3K9me3 and H3K27me3 in worm (observed with several validated antibodies; Extended Data Fig. 2b) suggests a species-specific difference in the organization of silent chromatin.

We also compared the patterns of histone modifications on expressed and silent genes in euchromatin and heterochromatin (Extended Data Fig. 2c and Supplementary Fig. 36). We previously reported prominent depletion of H3K9me3 at TSSs and high levels of H3K9me3 in the gene bodies of expressed genes located in fly heterochromatin¹⁴, and now find a similar pattern in human (Extended Data Fig. 2c and Supplementary Fig. 36). In these two species, H3K9me3 is highly enriched in the body of both expressed and silent genes in heterochromatic regions. In contrast, expressed genes in worm heterochromatin have lower H3K9me3 enrichment across gene bodies compared to silent genes (Extended Data Fig. 2c and Supplementary Figs 36, and 37). There are also conspicuous differences in the patterns of H3K27me3 in the three organisms. In human and fly, H3K27me3 is highly associated with silent genes in euchromatic regions, but not with silent genes in heterochromatic regions. In contrast, consistent with the worm-specific association between H3K27me3 and H3K9me3, we observe high levels of H3K27me3 on silent genes in worm heterochromatin, whereas silent euchromatic genes show modest enrichment of H3K27me3 (Extended Data Fig. 2c and Supplementary Fig. 36).

Our results suggest three distinct types of repressed chromatin (Extended Data Fig. 3). The first contains H3K27me3 with little or no H3K9me3 (human and fly states 10 and 11, and worm state 11), corresponding to developmentally regulated Polycomb-silenced domains in human and fly, and probably in worm as well. The second is enriched for H3K9me3 and lacks H3K27me3 (human and fly states 12 and 13), corresponding to constitutive, predominantly pericentric heterochromatin in human and fly, which is essentially absent from the worm genome. The third contains both H3K9me3 and H3K27me3 and occurs predominantly in worm (worm states 10, 12 and 13). Co-occurrence of these marks is consistent with the observation that H3K9me3 and H3K27me3 are both required for silencing of heterochromatic transgenes in worms¹⁶. H3K9me3 and H3K27me3 may reside on the same or adjacent nucleosomes in individual cells^{17,18}; alternatively the two marks may occur in different

marks. The genomic coverage of each state in each cell type or developmental stage is also shown (see Supplementary Figs 26–32 for detailed analysis of the states). States are named for putative functional characteristics.

cell types in the embryos and larvae analysed here. Further studies are needed to resolve this and determine the functional consequences of the overlapping distributions of H3K9me3 and H3K27me3 observed in worm.

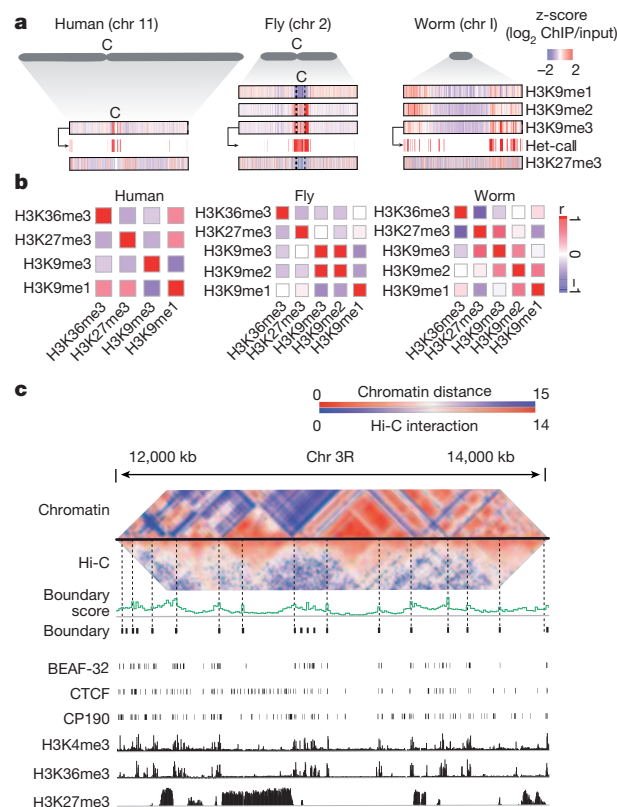


Figure 3 | Genome-wide organization of heterochromatin. **a**, Enrichment profiles of H3K9me1, H3K9me2, H3K9me3 and H3K27me3, and identification of heterochromatin domains based on H3K9me3 (illustrated for human H1-hESC, fly L3 and worm L3). For fly chr 2, 2L, 2LHet, 2RHet and 2R are concatenated (dashed lines). C, centromere; Het, heterochromatin. **b**, Genome-wide correlation among H3K9me1, H3K9me2, H3K9me3, H3K27me3 and H3K36me3 in human K562 cells, fly L3 and worm L3; no H3K9me2 profile is available for human. **c**, Comparison of Hi-C-based and chromatin-based topological domains in fly L3. Heat maps of similarity matrices for histone modification and Hi-C interaction frequencies are juxtaposed (see Supplementary Fig. 40).

Genome-wide chromatin conformation capture (Hi-C) assays have revealed prominent topological domains in human¹⁹ and fly^{20,21}. Although their boundaries are enriched for insulator elements and active genes^{19,20} (Supplementary Fig. 38), the interiors generally contain a relatively uniform chromatin state: active, Polycomb-repressed, heterochromatin, or low signal²² (Supplementary Fig. 39). We found that chromatin state similarity between neighbouring regions correlates with chromatin interaction domains determined by Hi-C (Fig. 3c, Supplementary Fig. 40 and Supplementary Methods). This suggests that topological domains can be largely predicted by chromatin marks when Hi-C data are not available (Supplementary Figs 41 and 42).

C. elegans and *D. melanogaster* have been used extensively for understanding human gene function, development and disease. Our analyses of chromatin architecture and the large public resource we have generated provide a blueprint for interpreting experimental results in these model systems, extending their relevance to human biology. They also provide a foundation for researchers to investigate how diverse genome functions are regulated in the context of chromatin structure.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 November 2013; accepted 29 April 2014.


- modENCODE Consortium *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
- Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Gassmann, R. *et al.* An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. *Nature* **484**, 534–537 (2012).
- Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002).
- Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
- Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* <http://dx.doi.org/10.1038/nature13424> (this issue).
- Boyle, A. P. *et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* <http://dx.doi.org/10.1038/nature13668> (this issue).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
- Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2011).
- Elgin, S. C. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5**, a017780 (2013).
- Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **21**, 227–236 (2011).
- Riddle, N. C. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.* **21**, 147–163 (2011).
- Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
- Towbin, B. D. *et al.* Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
- Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
- Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151**, 181–193 (2012).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
- Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
- Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was mainly funded by NHGRI U01HG004258 (G.H.K., S.C.R.E., M.I.K., P.J.P., V.P.), U01HG004270 (J.D.L., J.A., A.F.D., X.S.L., S.S.), U01HG004279 (D.M.M.), U54HG004570 (B.E.B.) and U01HG004695 (W.S.N.). It is also supported by NHBIB 5RL9EB008539 (J.W.K.H.), NHGRI K99HG006259 (M.M.H.), NIGMS fellowships (S.C.J.P., E.N.L.), NIH U54CA121852 (T.D.T.), NSF 1122374 (D.S.D.), National Natural Science Foundation of China 31028011 (X.S.L.), MEST Korea MHW-2013-HI13C2164 (J.H.K.), NRF-2012-0000994 (K.-A.S.), and Wellcome Trust 54523 (J.A.). We thank D. Acevedo and C. Kennedy for technical assistance.

Author Contributions J.W.K.H., Y.L.J., T.L., B.H.A., S.L., K.-A.S., M.Y.T., S.C.J.P., A.K., E.B., S.S.H. and A.R. led the data analysis. K.I., A.M., A.A., T.G., N.C.R., T.A.E., A.A.A. and D.A. led the data production. J.A.B., D.S.D., X.D., F.F., N.G., P.H., M.M.H., P.V.K., N.K., E.N.L., M.W.L., R.P., N.S., C.W. and H.X. analysed data. S.K.B., Q.B.C., R.A.-J.C., Y.D., A.C.D., C.B.E., S.E., J.M.G., D.H., M.H., T.E.J., P.K.-Z., C.V.K., S.A.L., I.L., X.L., H.N.P., A.P., B.Q., P.S., Y.B.S., A.V. and C.M.W. produced data. E.A.F., P.J.G. and M.J.P. carried out NIH scientific project management. The role of the NIH Project Management Group was limited to coordination and scientific management of the modENCODE and ENCODE consortia. J.W.K.H., Y.L.J., T.L., B.H.A., S.L., K.-A.S., M.Y.T., S.C.J.P., S.S.H., A.R., K.I., T.D.T., M.K., D.M.M., S.S., S.C.R.E., X.S.L., J.D.L., J.A., G.H.K. and P.J.P. wrote the paper. The group leaders for data analysis or production were R.E.K., J.H.K., B.E.B., A.F.D., V.P., M.I.K., W.S.N., T.D.T., M.K., D.M.M., S.S., S.C.R.E., J.A., X.S.L., G.H.K., J.D.L. and P.J.P. The overall project managers were D.M.M., S.S., S.C.R.E., X.S.L., J.D.L., J.A., G.H.K. and P.J.P.

Author Information All data are available at the project data portals (<http://data.modencode.org>; <http://encodeproject.org>), modMine (<http://intermine.modencode.org>), or our database and web application (http://encode-x.med.harvard.edu/data_sets/chromatin/). More information and annotated files are available at <http://encodeproject.org/comparative>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. B.H.A., S.L., K.I., K.-A.S., A.M., M.Y.T., A.A., S.C.J.P., T.G., A.K., N.C.R., E.B., T.A.E., S.S.H., A.A.A., A.R., D.A. contributed equally and should be considered co-second authors. Correspondence and requests for materials should be addressed to D.M.M. (david.macalpine@duke.edu), S.S. (sstrome@ucsc.edu), S.C.R.E. (selgin@biology2.wustl.edu), X.S.L. (xslu@jimmy.harvard.edu), J.D.L. (jdlieb@uchicago.edu), J.A. (ja219@cam.ac.uk), G.H.K. (ghkarpen@lbl.gov), or P.J.P. (peter_park@hms.harvard.edu). Questions about worm data should be directed to S.S., J.D.L. or J.A., fly data to D.M.M., S.C.R.E. or G.H.K., and analysis to X.S.L. and P.J.P.

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Comparative analysis of regulatory information and circuits across distant species

Alan P. Boyle^{1*}, Carlos L. Araya^{1*}, Cathleen Brdlik¹, Philip Cayting¹, Chao Cheng², Yong Cheng¹, Kathryn Gardner³, LaDeana W. Hillier⁴, Judith Janette³, Lixia Jiang¹, Dionna Kasper³, Trupti Kawli¹, Pouya Kheradpour⁶, Anshul Kundaje^{5,6}, Jingyi Jessica Li^{7,8}, Lijia Ma⁴, Wei Niu³, E. Jay Rehm⁹, Joel Rozowsky², Matthew Slattery⁹, Rebecca Spokony⁹, Robert Terrell⁴, Dionne Vafeados⁴, Daifeng Wang², Peter Weisdepp⁴, Yi-Chieh Wu⁶, Dan Xie¹, Koon-Kiu Yan², Elise A. Feingold¹⁰, Peter J. Good¹⁰, Michael J. Pazin¹⁰, Haiyan Huang⁷, Peter J. Bickel⁷, Steven E. Brenner^{11,12}, Valerie Reinke³, Robert H. Waterston⁴, Mark Gerstein², Kevin P. White^{9§}, Manolis Kellis^{6§} & Michael Snyder^{1§}

Despite the large evolutionary distances between metazoan species, they can show remarkable commonalities in their biology, and this has helped to establish fly and worm as model organisms for human biology^{1,2}. Although studies of individual elements and factors have explored similarities in gene regulation, a large-scale comparative analysis of basic principles of transcriptional regulatory features is lacking. Here we map the genome-wide binding locations of 165 human, 93 worm and 52 fly transcription regulatory factors, generating a total of 1,019 data sets from diverse cell types, developmental stages, or conditions in the three species, of which 498 (48.9%) are presented here for the first time. We find that structural properties of regulatory networks are remarkably conserved and that orthologous regulatory factor families recognize similar binding motifs *in vivo* and show some similar co-associations. Our results suggest that gene-regulatory properties previously observed for individual factors are general principles of metazoan regulation that are remarkably well-preserved despite extensive functional divergence of individual network connections. The comparative maps of regulatory circuitry provided here will drive an improved understanding of the regulatory underpinnings of model organism biology and how these relate to human biology, development and disease.

Transcription regulatory factors guide the development and cellular activities of all organisms through highly cooperative and dynamic control of gene expression programs. Regulatory factor coding genes are often conserved across deep phylogenies, their DNA-binding protein domains are preferentially conserved at the amino-acid level, and their *in vitro* binding specificities are also frequently conserved across large distances^{3,4}. However, the specific DNA targets and binding partners of regulators can evolve much more rapidly than DNA-binding domains, making it unclear whether the *in vivo* binding properties of regulatory factors are conserved across large evolutionary distances.

Comparisons of the locations of regulatory binding across species has been controversial, with some studies suggesting extensive conservation^{1,2,5–10}, whereas others suggest extensive turnover^{11–14}. Although it is generally assumed that across very large evolutionary distances regulatory circuitry is largely diverged, there exist highly conserved sub-networks^{15–18}. Thus, confusion exists in the level of regulatory turnover between related species, possibly owing to the small number of factors studied. Moreover, despite recent observations of the architecture of metazoan regulatory networks a direct comparison of their topology and structure—such as clustered binding and regulatory network motif—has not been possible owing to large differences in the procedures employed

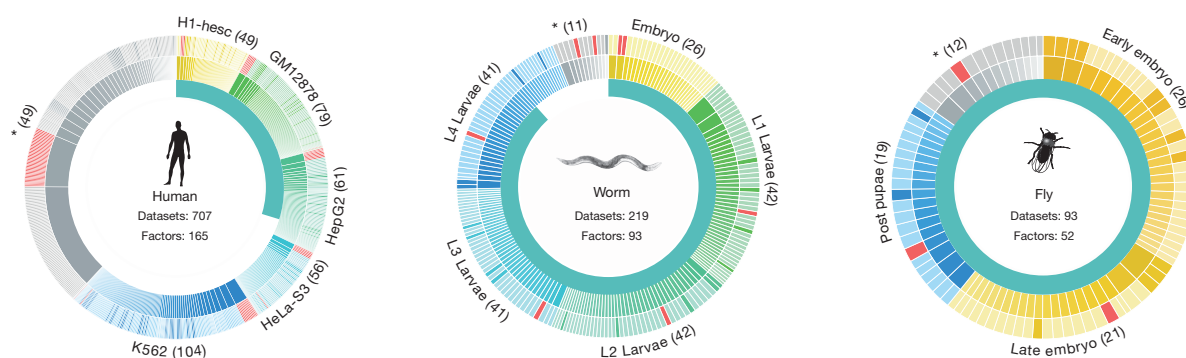


Figure 1 | A large compendium of regulatory binding across distant metazoans. Data from modENCODE and ENCODE consortia used in the analyses. Inner circles show the fraction of data sets presented here for the first time. For each organism the major contexts are shown in a different hue in the two outer circles. Asterisks, data sets that are not one of the main contexts. Each factor that underwent ChIP is shown in the middle circle and the count is

in parentheses (a factor can be represented in multiple contexts). The outer circle shows every data set, scaled by the number of peaks. Red, polymerase; light shades, transcription factor; dark shades, other. ChIP of a total of 165, 93 and 52 unique factors across all conditions and cell lines in human, and developmental stages in worm and fly, respectively.

¹Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ²Program of Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA. ³Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA. ⁴Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁵Department of Computer Science, Stanford University, Stanford, California 94305, USA. ⁶Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁷Department of Statistics, University of California, Berkeley, California 94720, USA. ⁸Department of Statistics, University of California, Los Angeles, California 90095, USA. ⁹Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. ¹⁰National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA. ¹¹Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ¹²Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

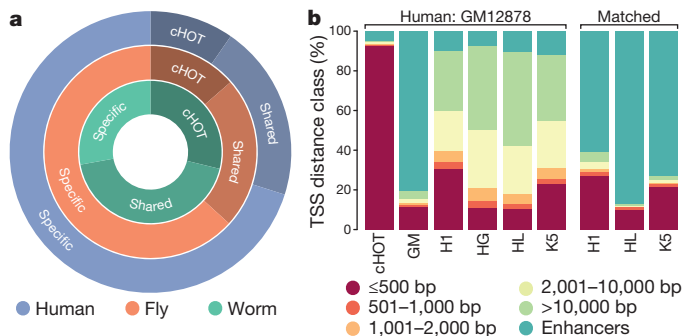


Figure 2 | HOTA regions are dynamically established across stages and cell types. HOTA regions contain binding sites for a large number of factors. **a**, A total of 2,948, 2,283, and 46,348 HOTA regions exist, of which 29.1%, 13.7% and 9.7% are constitutive in worm, fly and human respectively. A large fraction of HOTA regions are shared across multiple contexts but the majority of HOTA regions are specific to a single context. **b**, Constitutive human HOTA (cHOT) regions show strong enrichment for promoters while cell-type specific (GM12878 (GM), H1hesc (H1), HepG2 (HG), HeLaS3 (HL), K562 (K5)) HOTA regions show more enhancer enrichment (see also Extended Data Fig. 3). The cell type/context of the classes is indicated on top. Matched indicates that the classes are derived from the specific cell type analysed in each set.

to assay regulatory factor binding in distinct species. Here we present a systematic and uniform comparison of regulation using many factors across distantly related species to help address these questions on a scale not previously possible.

To compare regulatory architecture and binding across diverse organisms, the modENCODE and ENCODE consortia mapped the binding locations of 93 *Caenorhabditis elegans* regulatory factors, 52 *Drosophila melanogaster* regulatory factors and 165 human regulatory factors as a community resource (Fig. 1 and Supplementary Table 1). These regulatory factor binding data sets represent a substantial increase over those previously published for worm (194 new data sets for a total of 219) and

human (211 new, 707 total) and a substantial improvement in data quality in fly with a move from chromatin immunoprecipitation with DNA microarray (ChIP-chip) to ChIP followed by sequencing (ChIP-seq) (93 new, 93 total)^{2,8,19,20}. The majority of regulatory factors are site-specific transcription factors (83 in worm, 41 in fly, and 119 in human), although general regulatory factors such as RNA Pol II were also assayed.

All regulatory factors were analysed by ChIP-seq according to modENCODE/ENCODE standards: antibodies were extensively characterized, and at least two independent biological replicates were analysed²¹. Worm regulatory factors were assayed in embryo and stage 1–4 larvae (L1–L4 larvae), fly regulatory factors in early embryo, late embryo and post embryo, and human regulatory factors in myelocytic leukaemia K562 cells, lymphoblastoid GM12878 cells, H1 embryonic stem cells, cervical cancer HeLa cells, and liver epithelium HepG2 cells. Binding sites were scored using a uniform pipeline that identifies reproducible targets using irreproducible discovery rate (IDR) analysis (Extended Data Fig. 1)²² and quality-filtered experiments (see Methods). These rigorous quality metrics insure that the data sets used here are robust. All data presented are available at <http://www.ENCODEProject.org/comparative/regulation/>.

To explore motif conservation, we examined the 31 cases in which we had members of orthologous transcription-factor families profiled in at least two species (Extended Data Fig. 2a and Methods). Sequence enriched motifs were found for 18 of the 31 families and for 12 orthologous families (41 regulatory factors), the same motif is enriched in both species (Extended Data Fig. 2b, c). For 18 of 31 families (64 of 93 regulatory factors), the motif from one species is enriched in the bound regions of another species (one-sided hypergeometric, $P = 3.3 \times 10^{-4}$). These findings indicate that many factors retain highly similar *in vivo* sequence specificity within orthologous families, a feature noted previously across studies working on smaller numbers of factors.

Next, we used RNA-seq data³ to determine whether targets of orthologous regulatory factors are specifically expressed at similar developmental stages between fly and worm. As a class, orthologous regulatory factors (both assayed here and not) are significantly expressed at similar

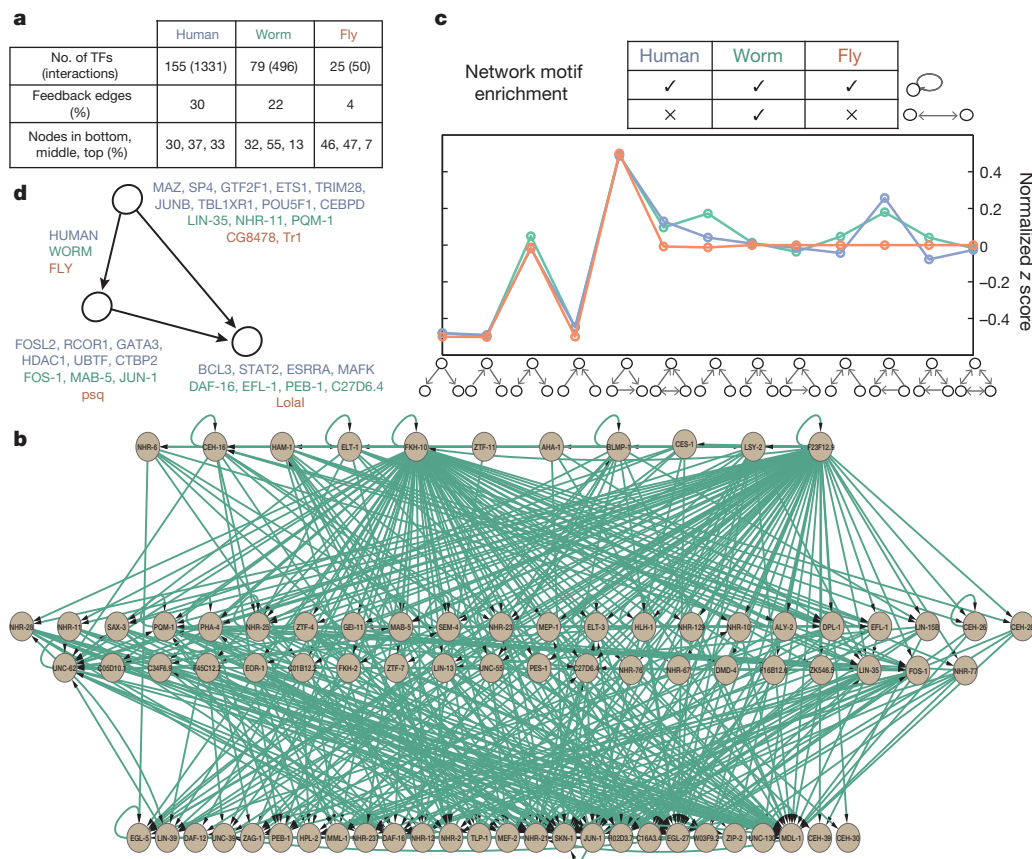


Figure 3 | Regulatory network motif usage is conserved across distant metazoans. **a**, Statistics of the transcription regulatory networks in human, worm, fly and their hierarchical organization. **b**, An example of the hierarchical network for worm. **c**, Network motif enrichment. The human, worm and fly networks are mostly consistent in terms of motif enrichment. The motif feed-forward loop is the most enriched motif in all three networks. **d**, Different transcription factors have different tendencies to appear as top, middle and bottom regulators in a FFL. The lists of human, worm, fly transcription factors with corresponding tendencies are displayed.

stages (Extended Data Fig. 3a–c). However, expression of orthologous targets of orthologous regulatory factors in worm and fly shows little significant target overlap (Extended Data Fig. 3d) and the large majority of orthologous regulatory factors did not show conserved target functions (Extended Data Fig. 4a–c), suggesting extensive re-wiring of regulatory control across metazoans. Nevertheless, human and worm orthologous regulatory factors were more likely to show conserved target gene functions than non-orthologous regulatory factors (Extended Data Fig. 4d, Wilcoxon test $P < 3.9 \times 10^{-6}$), highlighting regulatory factors with conserved target functions.

Regulatory factor binding is not randomly distributed throughout the genome, but rather, in all three species, approximately 50% of binding events are found in highly-occupied clusters, termed high-occupancy target (HOT) regions^{1,2,5,8,10}. HOT regions show enhancer function in integrated transcriptional reporters¹¹ and are stabilized by cohesin^{15,17}. HOT regions show no significant enrichment with non-specific antibodies (Extended Data Fig. 5), in contrast to recent work using raw signal¹⁹ rather than IDR peaks, although the possibility that they are artefacts has been raised.

By comparing HOT regions across different developmental times and cells types, we find that 5–10% of HOT regions are constitutive, indicating that HOT regions are dynamically established, rather than an intrinsic property of specific regions. In humans we find that approximately 90% of constitutive HOT regions fall within promoter chromatin states compared to only approximately 10–20% of context-specific HOT regions (Fig. 2a and Extended Data Fig. 6). Instead, approximately 80–90% of context-specific HOT regions fall within enhancer states. Moreover, these context-specific HOT regions are specifically enriched for enhancers in matching cell types or developmental stages. For example, 80% of GM12878-called HOT regions fall within GM12878-specific enhancers but only approximately 10% of GM12878-called HOT regions

fall within enhancers called in other cell-types (Fig. 2b). These patterns remain similar for all cell types (Extended Data Fig. 7), suggesting the two types of HOT regions are established concordantly and dynamically between cell types, though these patterns are weaker in the worm and fly data.

We constructed regulatory networks in each species by predicting gene targets of each regulatory factor using TIP²³ and used simulated annealing to reveal the organization of regulatory factors in three layers of master-regulators, intermediate regulators, and low-level regulators (Fig. 3a, b). The algorithm found only 7% of regulatory factors at the top layer of the network in fly and 13% in worm, compared to 33% in human. We also found that more edges are upward flowing in human (30%) than worm and fly (22% and 7%). This suggests differences in the global network organization with more extensive feedback and a higher number of master regulators in human.

We next assessed the local structure of regulatory networks, by searching for enriched sub-graphs known as network motifs (Fig. 3c). We found that the same network motifs were most and least enriched in the three species. In each case, the most abundant was the feed-forward loop (FFL), while the least abundant were cascade motifs, and both divergent and convergent regulation. Moreover, specific regulatory factors were enriched for origin, target, or intermediate regulators in these FFLs in each species (Fig. 3d). Surprisingly, the number of feed forward loops (FFLs) varied by developmental stage in both worm and fly, with L1 stage in worm and late-embryo stage in fly showing the highest number of FFLs (Extended Data Fig. 8), suggesting increased filtering fluctuations and accelerating responses in these stages²⁴.

We asked whether the three species showed conserved regulatory factor co-associations. We first focused on global co-associations where two factors co-associate frequently regardless of context, either by intermolecular interactions or independent recruitment (Extended Data

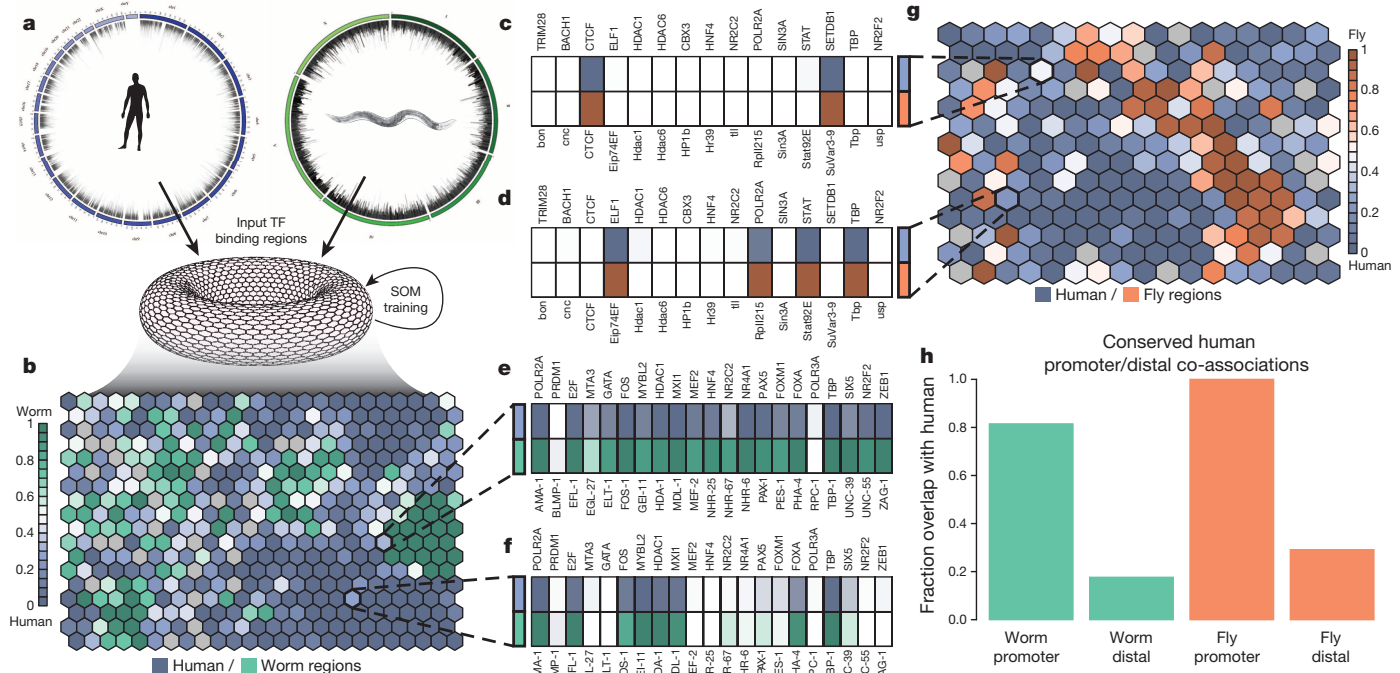


Figure 4 | Transcription-factor co-association. Many instances of transcription-factor co-association are under very specific contexts and probably not observed in a simple genome-wide co-association study. **a**, We combined the patterns of orthologous factors and genomic regions from two organisms to train a SOM where each 'hexagon' contains genomic regions from either organism with the same binding pattern of orthologous factors for worm (**b**) and fly (**g**). Each hexagon is shaded by the frequency of the pattern in the pairs of organisms. We show an example of binding patterns of 4 hexagons from the human–fly (**c–d**) and the human–worm (**e–f**). Names above the heatmaps are human factor names, and those below are their orthologue

names. Dark shaded boxes indicate binding of that factor. **c**, A binding pattern shared at equal frequency between human and fly with only CTCF and SETDB1 (CTCF and SuVar3-9 in fly) binding. **d**, A binding pattern that occurs more frequently in human shows ELF1, RNA Pol II, STAT and TBP binding. **e**, A binding pattern at similar frequencies in human and worm that is an example of a HOT region. **f**, A pattern more frequent in humans than worms shows RNA Pol II, E2F, FOS, MYBL2, HDAC1, MXI1, FOXA and TBP binding. **h**, Co-localization patterns that occur more frequently near promoters (<500 bp) in humans are highly likely to also occur at promoters in worm (80%) and fly (100%).

Fig. 9). With the exception of a small number of conserved global regulatory factor co-associations (for example, SIN3A with HDAC1, HDAC2 and NR2C2 in fly and human^{25–27}, and MXI1 with E2F1, E2F4 and E2F6 in worm and human), the majority of global co-associations were not conserved in the contexts and species pairs analyzed.

As regulatory factor co-association at distinct binding regions is local and contextual (that is, different combinations of factors co-associate at different genomic locations), we next used an approach to detect co-association at distinct regions of the genome based on conserved patterns of regulatory factor binding. This method uses self-organizing maps (SOMs) to analyse co-association patterns at specific loci by better exploring the full combinatorial space of regulatory factor binding than traditional co-association approaches (Fig. 4a–c)²⁸. We demonstrate that co-associations at distinct genomic regions reveal a more complex view of regulatory structure and bring forth categorical enrichments that are lost in a larger, genomic context.

We examined whether specific contextual co-associations are conserved for orthologous regulatory factors by using binding data from each organismal pair; that is, human–worm and human–fly (Fig. 4b, g). Specific regulatory factor co-associations were observed; most are conserved to varying degrees across each organism with very few that are entirely organism-specific (Fig. 4b, g). These co-associations result in expected sets of factors such as the previously noted SIN3A + HDAC co-association. In addition, we find new co-associations such as the pattern in Fig. 4f for human–worm, which in worm is highly enriched for GO terms associated with sex determination. We further examined which co-associations are conserved at distinct gene locations (that is, proximal and distal). We found distinct combinations of conserved co-associations in relation to transcription start site (TSS) regions. Interestingly, virtually all TSS-proximal co-associations in human remain TSS-proximal in worm (approximately 80%) and fly (approximately 100%), indicating that co-associations that occur at promoters are often highly conserved (Fig. 4h). Conversely, co-associations at distal regions are much less conserved.

Our results, obtained using a large resource of regulatory binding information, suggest that there is little conservation of individual regulatory targets and binding patterns for these highly divergent metazoans: *C. elegans*, *D. melanogaster* and *H. sapiens*. However, we do find strong conservation of overall regulatory architecture, both in network motif usage and in concentrated regulatory binding at dynamically established HOT regions. We observe an increased conservation of *in vivo* sequence preferences and some target gene functions, with context-specific regulatory factor partners still observed at specific loci in these distal comparisons. These findings are consistent with previous results indicating that the gene targets of regulation are typically quite divergent and are likely to account for many of the phenotypic differences among species^{12–14,16,29,30}, despite conserved sequence preferences. We significantly extend these observations, both in the number of regulators studied and in the range of regulatory properties studied, and provide specific examples of conserved and diverged regulatory functions. Lastly, beyond its potential for comparative studies of gene regulation, the primary data sets provide invaluable new information of genome-wide transcription-factor binding information both in human, and in two of the most important metazoan models of human biology, development, and disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 November 2013; accepted 10 July 2014.

1. modENCODE Consortium et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
2. Gerstein, M. B. et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
3. Gerstein, M. et al. Comparative analysis of the transcriptome across distant species. *Nature* <http://dx.doi.org/10.1038/nature13424> (this issue).
4. Berger, M. F. et al. Variation in homeodomain dna binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).

5. Moorman, C. et al. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
6. Lavoie, H. et al. Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* **8**, e1000329 (2010).
7. He, Q. et al. High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nature Genet.* **43**, 414–420 (2011).
8. ENCODE Project Consortium et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
9. Mikkelsen, T. S. et al. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156–169 (2010).
10. Yip, K. Y. et al. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13**, R48 (2012).
11. Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J. & Stark, A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev.* **26**, 908–913 (2012).
12. Schmidt, D. et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
13. Odom, D. T. et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39**, 730–732 (2007).
14. Borneman, A. R. et al. Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
15. Yan, J. et al. Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801–813 (2013).
16. Peter, I. S. & Davidson, E. H. Evolution of gene regulatory networks controlling body plan development. *Cell* **144**, 970–985 (2011).
17. Faure, A. J. et al. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* **22**, 2163–2175 (2012).
18. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Rev. Genet.* **13**, 613–626 (2012).
19. Teytelman, L., Thurtle, D. M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl Acad. Sci. USA* **110**, 18602–18607 (2013).
20. Nègre, N. et al. A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
21. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
22. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
23. Cheng, C., Min, R. & Gerstein, M. TIP: a probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles. *Bioinformatics* **27**, 3221–3227 (2011).
24. Alon, U. Network motifs: theory and experimental approaches. *Nature Rev. Genet.* **8**, 450–461 (2007).
25. Heinzel, T. et al. A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* **387**, 43–48 (1997).
26. Nan, X. et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
27. Huang, Y., Myers, S. J. & Dingleline, R. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nature Neurosci.* **2**, 867–872 (1999).
28. Xie, D. et al. Dynamic trans-acting factor colocalization in human cells. *Cell* **155**, 713–724 (2013).
29. Carroll, S. B., Grenier, J. & Weatherbee, S. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Wiley-Blackwell, 2004).
30. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is supported by the NHGRI as part of the modENCODE and ENCODE projects. This work was funded by U01HG004264, RC2HG005679 and P50GM081892 to K.P.W., U54HG006996, U54HG004558 and U01HG004267 to M.S., and F32GM101778 to K.E.G.

Author Contributions A.P.B., C.L.A., Y.C., D.X., P.K., A.K., P.C., L.M., K.K.Y., J.R., D.W., C.C., L.H., P.C. and Y.C.W. were involved in data analysis. M.S., R.S., E.J.R., D.V., R.T., P.W., R.H.W., C.B., K.G., J.J., L.J., D.K., T.K., W.N. and R.S. produced data. A.P.B., M.S., C.L.A., K.W., K.K.Y. and R.H.W. wrote the paper. E.A.F., P.J.G. and M.J.P. carried out NIH scientific project management. The role of the NIH Project Management Group in the preparation of this paper was limited to coordination and scientific management of the modENCODE and ENCODE consortia. M.S., M.K., K.P.W., M.G., R.H.W. and V.R. were responsible for overall project management.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (mnsnyder@stanford.edu), M.K. (manoli@mit.edu) or K.P.W. (kpwhite@uchicago.edu).



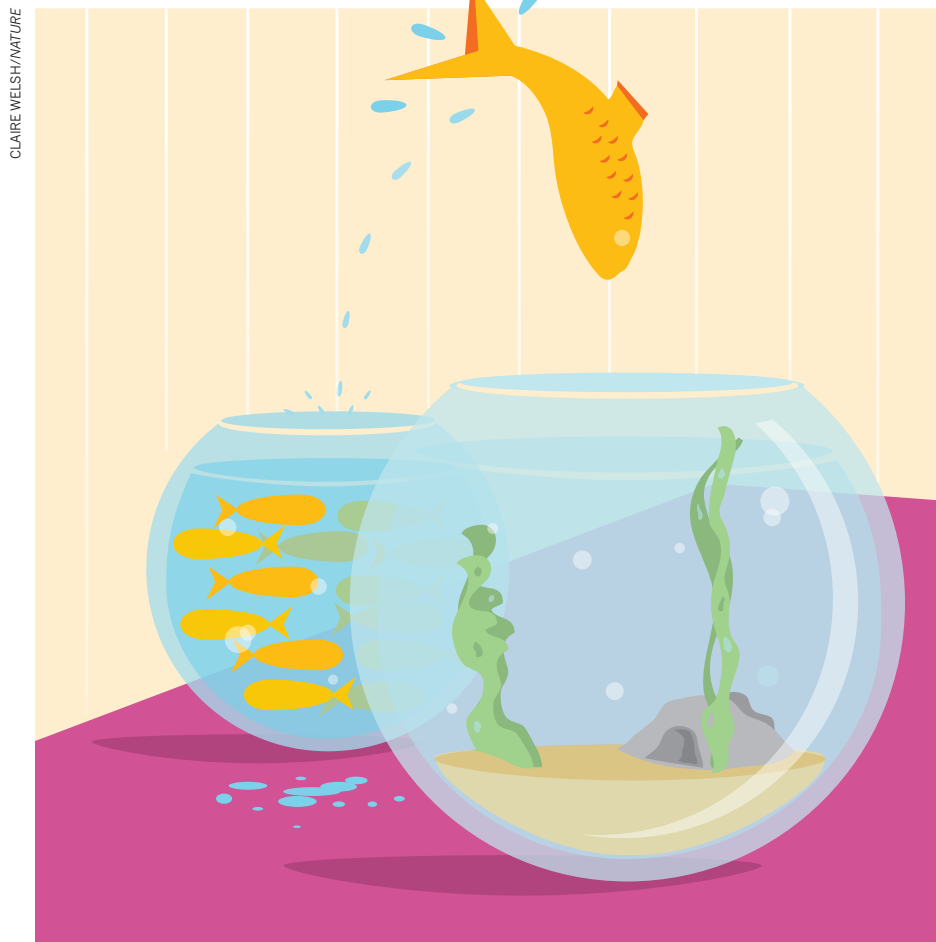
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

CAREERS

FUNDING Growth slips in US charitable giving to higher education **p.458**

JOB APPLICATIONS Falsehoods on CVs can jeopardize a job offer **p.458**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



CLAIRE WELSH/NATURE

began grooming him, and the other males, even the largest alphas, gave him a wide berth.

In one afternoon, Mike went from one of the lowest-status members of the group to one of the highest. He had adopted a strategy known in job-seeking circles as the bold move.

A bold move is a calculated strategy to draw attention to yourself through a dramatic and unusual act. It signals to a potential employer that you have creativity, drive and the willingness to take risks to achieve something important. When executed convincingly and with sincerity, the bold move can completely reset an employer's perceptions of your abilities and potential, because in the world outside academia, the bold move can get you ahead.

TICK TOCK

In a talk last year to PhD students at the University of California, Berkeley, author and lecturer Kristina Susac told of a bold-move strategy she had successfully used in a former sales post to nab a face-to-face sales meeting with an evasive prospect, the chief information officer of a large company. After many failed attempts to reach the executive through phone calls and e-mails, Susac realized that these methods were not going to produce a meeting.

So she sent the executive a box by courier. Inside the box was a large alarm clock with a tag that read, "Isn't it time that we met?" Susac's phone rang the next morning: it was the executive. "OK, I give in — that was one of the most unusual and creative stunts that I have seen from any salesperson," the officer told her. "Why don't you come in next week and we can discuss what you have to offer."

I told this story to a graduate student who was keenly interested in getting an internship as a copywriter in an advertising firm. Although he was an excellent and creative writer and editor, he lacked the copywriting experience that was listed as a non-negotiable requirement in all the job advertisements that he had seen. So after a brief informational interview with a member of the firm, he designed a 'campaign' T-shirt that bore an image of his face in profile, his name and the words "Running for Intern, 2014". He sent the T-shirts to the firm's vice-president and head of creative products, and persuaded several other employees whom he had met during his interview to wear the T-shirts. He got the post.

Early-career scientists often cringe when they first hear about the bold-move strategy. After spending years at the bottom of the academic hierarchy, many young researchers ►

COLUMN

Bold moves

People seeking non-academic jobs may need to try something unexpected to be noticed, says **Peter Fiske**.

In Jane Goodall's landmark study of chimpanzees in Gombe, Tanzania, she observed one young male who devised an unusual means of raising his status. Dubbed Mike by Goodall's researchers, the youngster had been bullied by higher-status males, rarely received portions of food from others and was locked out of any potential for procreation (H. Kummer and J. Goodall *Phil. Trans. R. Soc. Lond. B* **308**, 203–214; 1985).

But one day Mike discovered that when he accidentally knocked against the large metal kerosene cans lying around Goodall's camp, the racket startled the other members of the chimpanzee group. One afternoon, Mike ran into the centre of the group, batting and banging on one of the cans and scattering all the chimpanzees. Then he sat quietly in the centre of the clearing until the other chimpanzees slowly returned to gather around him. The females

FUNDING

Donation expectations

US higher-education institutions reported that the value of charitable donations they received grew 5.2% for the 2013–14 academic year, according to survey results from the Council for Advancement and Support of Education (CASE) in Washington DC. The institutions also predicted a growth of 5.7% for 2014–15. Community colleges, which offer basic tertiary education, reported a 7.2% growth in giving for 2013–14, compared with 5% for public and private universities that offer full, four-year degrees. The average yearly growth over 20 years is 5.9%. CASE polls institutions twice a year on actual and predicted giving.

TEACHING

Mentoring programme

The New York Academy of Sciences and Sister Cities International in Washington DC are collaborating on a mentoring programme in science, technology, engineering and maths (STEM) that will match graduate and undergraduate students with STEM professionals. The STEMentoring Program will draw mentors and students from Sister City volunteers around the world and match them through an online platform. Students will be able to connect with their counterparts and professionals globally through a virtual network, in which they can also participate in seminars and complete coursework. The programme will also foster collaboration in real-world STEM-related issues, such as water availability, in communities worldwide.

JOB APPLICATIONS

Lies have consequences

Early-career researchers looking for positions in industry can probably wave goodbye to a job if they lie on their CVs, finds a survey of US hiring managers and human-resources executives. Conducted online from 13 May to 6 June by Harris Poll in Rochester, New York, on behalf of the jobs website CareerBuilder, the survey polled 2,188 respondents across various industries and company sizes. Some 51% of respondents said that they would dismiss an applicant on whose CV they found a lie, whereas 7% said that they would overlook it if they liked the candidate. Respondents said that 33% of detected untruths involved an applicant's academic degree, and 57% were used to embellish skill sets.



KENNAN WARD/CORBIS

Primatologist Jane Goodall observed a young chimpanzee making a bold move to earn social status.

believe that their careers will progress only if they patiently work their way up the ladder, with advancement coming as the 'silverbacks' — people reaching retirement age — retreat. Sudden elevations in status are virtually unheard of in the slow-moving, heavily circumscribed academic universe, even after publication of a world-class discovery.

In the academic culture, graduate students learn that to qualify for positions in their disciplines, they must be widely recognized experts, and that the only positions that they should consider are those closely aligned with their areas of research. The academic establishment views self-promotion with deep suspicion: the work should speak for itself. If graduate students try to draw attention to themselves or their work, some faculty members and department heads might see those efforts as a signal that the work is sub-par.

Although these cultural norms may be de rigueur in academia, they leave early-career researchers maladapted to the job market outside it. Expertise is the foremost qualification for a job inside academia, but in the outside world, attitude is at least as important to potential employers. In academia, a researcher can spend years focusing on a single problem or technical area. But in other, profit-driven sectors, employers know that priorities, opportunities and technologies change quickly. As an employee, adaptability and willingness to learn are more important to career success than is technical expertise.

If a young PhD-holder is modest about his or her talents when applying for non-academic jobs, it can reinforce the perception

among potential employers that he or she is uncertain, hesitant about taking the initiative and poorly suited to any job other than academic research — for which, of course, few open positions exist.

A bold move is, of course, inherently risky: a potential employer may view your attempt to set yourself apart as inappropriate. There is no way to know for certain whether your bold move is on target, but discussing it with a few friends or a mentor may help you to craft a gesture that is distinctive without being bizarre. As a science-trained PhD-holder, if you seek employment outside academia — and greater economic reward for your years of investment in graduate school — you need to be able to retain the qualities of honesty and intellectual integrity while adapting to the cultural norms of the non-academic working world. Understand that your graduate-school experience has given you a broad set of transferable skills that, when combined with your intelligence and resourcefulness, will enable you to succeed in a wide range of jobs and roles.

Some amount of self-advocacy is essential. If you do not assert what you believe you are capable of doing — and project a positive, confident attitude — employers will not take the risk of overlooking your lack of experience. Sometimes it is important to make that bold move and to take what might feel like a risky step. Young PhD researchers tend to be risk-averse, but what may seem like a risky move to them is probably not that risky at all. The occasional bold move sends the signal to potential employers that you are more than just your PhD — and that you are ready to move up and beyond the rest of the pack. ■

Peter Fiske is chief executive of PAX Water Technologies in Richmond, California, and author of Put Your Science to Work.

THE ANGLE OF THE LIGHT ON THE BLOODSTAINED KITCHEN FLOOR

4 p.m. Friday 22 November 1963.

BY MATT MIKALATOS

Doctor Wallace told me that the 22 November 1963 would be my assigned topic for the final exam. I've studied and crammed and memorized and watched holos and paid for memory boosts. But when I put on the Chronovisor, it doesn't show me Dallas, it shows me Tony Acerbi's apartment.

So I do the research while my exam timer counts down. Dr Wallace has given me a two-week window of choral study, culminating here on the 22nd.

I learn that Tony Acerbi lives in a fetid hole of an apartment. He has way more cockroaches than friends. He wears slacks and a filthy undershirt most of the day, and lives on boosted merchandise and odd jobs. His home is in a Chicago neighbourhood where you call the cops and make bets on whether they'll show. 60–40 they don't.

Tony's neighbour is this old black man named Ollie Lewis. He's tied with Tony for most hated man in the complex. Everyone's pretty sure he poisoned the barking dog in 3A, and he definitely empties his ashtray into Mrs Fortuna's flower box.

At 3.42 p.m. on 22 November, two teenagers break into Ollie's place.

Tony is sitting in his easy chair smoking a cigarette and he hears the muffled sounds of the old man next door getting worked over. He jumps up and, without putting on a jacket, or a shirt for that matter, he runs through the November cold and busts into Ollie's apartment.

Ollie is on the floor, curled in a ball. One kid has a Louisville Slugger, the other has steel-toed boots. Tony takes a couple of swings at them, but they're not interested. They keep taking jabs at Ollie. So Tony lays down over Ollie, covering the man's head with his own body. The kids barely pause, but now most

of their blows are hitting Tony. There is blood everywhere. Tony, at a certain point, just turns his head to one side and stares at the floor, and the fading sunlight through the dirty window illuminates every mote of dust, like a thousand bursting stars drifting through his neighbour's kitchen.



I can't go much beyond this moment in the kitchen, and I can't see places other than these two apartments. I can't figure out why the teens are doing this, or why Tony is protecting Ollie. I don't see a reason for him to lay himself down like a shroud over his neighbour.

I click my teeth in annoyance that Dr Wallace has done this to me, given me this particular moment on this day of all days. I have to wonder if Tony knows how inconsequential this moment is. A couple of hours from now, JFK takes a bullet to the head. Aldous Huxley and C. S. Lewis both breathe their last. Hell, today The Beatles release their second album. *Doctor Who* has its series premiere less than 24 hours from this moment.

Two men in a dilapidated apartment complex being beaten senseless by teenagers, it means less than nothing. You don't get grants, you don't get book deals, you don't get tenure by studying unimportant events.

The two men breathe shallowly and twitch on the yellowed linoleum. Their breath stirs ripples in the growing puddle of blood. The teenagers leave them, leaving the door open.

Two and a half minutes later, Mrs Fortuna

happens by to water her flowers and when she sees the men collapsed in Ollie's kitchen, she screams and runs for her phone and calls the police, the fire department, the ambulances, anyone, anyone, please God, just send anyone.

I probably couldn't even write a thesis about Tony Acerbi or Ollie Lewis. There

won't be a holo. They're nobodies and the resources for choral viewing are tight. Dr Wallace picked an obscure moment because he likes his students to observe details, to see things fresh, and it's hard to do that with the big historical events. So he chose this little corner of nowhere as a test, not because he thinks it's important.

What happened to these two doesn't matter. Not to my grade, not to my future, not to history or the fate of the country.

None of it matters. The identity of the teenagers, the amount of blood lost, the pattern in the wallpaper, the reason for the assault, the brand of Tony's cigarettes, the angle of the light that illuminates the two men gasping for breath on the kitchen floor. It doesn't matter, not at all, not one bit.

Out the broken doorway I can see the windows of a hundred other apartments, a hundred other stories that matter just as little as this one.

Still.

I find myself hoping for wailing sirens and blue and red lights. I find myself praying that someone will come.

And just like that, my time's up. Dr Wallace removes my Chronovisor and places it on the next student in line.

"What happened?" I ask.

He turns, a weary pity in his eyes and says: "You passed." ■

Matt Mikalatos lives near Portland, Oregon. He has a podcast called the *StoryMen* (www.storymen.us), a website called *Norville Rogers* (www.norvillerothers.com) and an inability to time travel in any direction but towards the future.

JACEY

➔ NATURE.COM

Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm